

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-222380

(43)Date of publication of application : 17.08.2001

(51)Int.Cl. G06F 3/06

G11B 19/02

G11B 20/10

(21)Application number : 2000-029286 (71)Applicant : HITACHI LTD

(22)Date of filing : 07.02.2000 (72)Inventor : TSUNODA MOTOYASU

HIRATSUKA YUKIE

IGUCHI SHINYA

NISHIKAWA MANABU

(54) EXTERNAL STORAGE DEVICE AND INFORMATION PROCESSING SYSTEM
WITH THE SAME

(57)Abstract:

PROBLEM TO BE SOLVED: To optimize an information processing system including an external storage device such as a hard disk device.

SOLUTION: A performance parameter (such as spindle rotating speed, cache size) to determine performance of a hard disk is transmitted and received between the hard disk device 101 and a host device 113. The hard disk device 101 resets the indicated performance parameter when a setting request of the performance parameter is received from the host device 113. In addition, the host device 113 predicts execution

time of commands to the hard disk device 101 based on parameter information received from the hard disk device 101 and rearranges the commands based on a prediction result.

LEGAL STATUS [Date of request for examination]
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.*** shows the word which can not be translated.

3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] External storage which is the external storage which an information processor accesses and is characterized by having a means to transmit and receive the engine-performance information which determines the self engine performance between said information processors, and a means to change the value of the engine-performance information on self for modification of the value of said engine-performance information in response to a demand from said information processor.

[Claim 2] Information processing system characterized by having external storage according to claim 1 and the information processor accessed to the external storage concerned.

[Claim 3] A means to receive the engine-performance information which is the information processor which accesses external storage and determines the engine performance of said external storage to the external storage concerned,

A command queue means to store two or more commands to said external storage, A means to predict command execution time amount about each command stored in said command queue means based on the engine-performance information received from said external storage, The information processor characterized by having the means which rearranges the execution sequence of the command stored in the command queue means so that a system might be optimized based on the predicted command execution time amount.

[Claim 4] The information processor according to claim 3 characterized by having further a means to receive the command to said external storage and to store in said command queue means from other information processors which access said external storage.

[Claim 5] Information processing system characterized by having an information processor according to claim 3 or 4 and the external storage which transmits said engine-performance information to said information processor.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to optimization of the information processing system containing external storage, such as a hard disk drive unit.

[0002]

[Description of the Prior Art] In recent years, it follows on large-capacity-izing of a

hard disk drive unit, and improvement in the speed, and the system which records / reproduces an animation and voice at a hard disk drive unit is proposed. As for an animation or data called voice, a real-time operation, i.e., the band guarantee of data transfer, is needed.

[0003] Therefore, the execution time required for this command processing was predicted, when the result was in the maximum allowed time which is demanding the host computer, the command was executed, when the I/O command which performs execution-time prediction was received from the host computer, when it seemed that allowed time was exceeded, activation of a command was stopped by the magnetic disk drive indicated by JP,10-222310,A, for example, and that was notified to the host computer.

[0004]

[Problem(s) to be Solved by the Invention] The technique given [above-mentioned] in an official report has put the chief aim on the band guarantee of data transfer, and only predicts command execution time amount from an internal state in itself [disk unit]. Therefore, the disk unit is not necessarily working in the optimal condition in a system. For example, when reproducing voice, supposing the transfer band of a disk unit is fully guaranteed,

it may be desirable to reduce the noise sound which reduces spindle rotational speed and a disk unit emits.

[0005] Moreover, with a technique given [above-mentioned] in an official report, when an I/O command cannot perform in allowed time, that is notified to a host computer, but the result is not obtained unless, as for a host computer, it actually publishes a command. Therefore, the case where a useless command is published to a disk unit arises. Moreover, when it sees by the whole system, the command is not necessarily published to the optimal timing to the disk unit.

[0006] The purpose of this invention is to offer a technique for the information processing system containing external storage, such as a hard disk drive unit, to work according to liking of a user, so that activation of each application may be optimized.

[0007] Moreover, other purposes of this invention are to offer the technique which optimizes issue of the command to external storage in order to raise the performance of a system.

[0008]

[Means for Solving the Problem] The external storage concerning this invention is external storage which an information processor accesses. And it is

characterized by having a means to transmit and receive the engine-performance information which determines the self engine performance between said information processors, and a means to change the value of the engine-performance information on self for modification of the value of said engine-performance information in response to a demand from said information processor.

[0009] Moreover, information processing system concerning this invention is characterized by having said external storage and the information processor accessed to the external storage concerned. In this case, an information processor requires modification of engine-performance information at the time of initialization of a system, and application activation.

[0010] Moreover, the information processor concerning this invention is an information processor which accesses external storage. And a means to receive the engine-performance information which determines the engine performance of said external storage to the external storage concerned, A command queue means to store two or more commands to said external storage, A means to predict command execution time amount about each command stored in said command queue means based on the engine-performance information received

from said external storage, It is characterized by having the means which rearranges the execution sequence of the command stored in the command queue means so that a system might be optimized based on the predicted command execution time amount.

[0011] In this case, the command to said external storage is received and you may make it have further a means to store in said command queue means, from other information processors which access said external storage.

[0012] Moreover, the command to said external storage has the flag information which shows that the band guarantee of data transfer is required, and you may make it rearrange the execution sequence of a command using the flag information concerned. Furthermore, as a result of predicting command execution time amount, when data transfer cannot be band guaranteed about a command with said flag information, activation of the command concerned is stopped and you may make it notify a user of that.

[0013] Moreover, an information processor is equipped with a means to measure an effective data transfer rate from a data transfer time, and you may make it predict the execution time of a command to be the amount of the data transfer produced with the command published to external storage to said external

storage or information processor using a measurement result.

[0014] Moreover, information processing system concerning this invention is characterized by having said information processor and the external storage which transmits said engine-performance information to said information processor.

[0015] In addition, when external storage is a disk unit, as said engine-performance information, spindle rotational speed, the cache control mode, cache size, the number of the maximum read-ahead sectors, a seek-time formula, sector-address transformation, the number of reserve sectors, ECC length, the number of on-the-fly ECC corrections, a host transfer rate, etc. are mentioned, for example.

[0016]

[Embodiment of the Invention] Hereafter, the gestalt of operation of this invention is explained to a detail, referring to a drawing.

<< -- 1st operation gestalt>> -- the 1st operation gestalt of this invention is explained first.

[0017] Drawing 1 is drawing showing the configuration of the information processing system which applied this invention. As shown in this drawing, this

information processing system consists of host equipment 113 (for example, personal computer) and the hard disk drive unit 101 as external storage.

[0018] With this operation gestalt, as an interface between host equipment 113 and a hard disk drive unit 101, although ATA (AT attachment) is assumed, other interfaces (for example, IEEE1394, SSA (serial storage architecture), Fiber Channel, etc.) can be adapted.

[0019] Host equipment 113 is equipped with the host CPU 114, a bus controller 115, and a host RAM 116 and the host interface controller 117. In addition, you may make it have the component with other host equipment 113 with a natural thing, for example, a graphics controller, a network controller, a floppy disk controller, etc.

[0020] The host CPU 114 performs predetermined processing by performing the program stored in ROM (un-illustrating) or the host RAM 116. A bus controller 115 controls the data transfer between the host CPU 114, a host RAM 116, and the host interface controller 117. The host interface controller 117 controls the data transfer between a hard disk drive unit 101 and a host RAM 116 etc.

[0021] On the other hand, a hard disk drive unit 101 is equipped with a disk CPU 102, the disk interfacing controller 103, a disk controller 104, a disk RAM 105

and the signal-processing section 106, the servo controller 107, and the HDA section 108.

[0022] A disk CPU 102 controls data processing of the hard disk drive unit 101 whole. A disk controller 104 controls the data transfer between the disk interfacing controller 103 and a disk RAM 105 and between the signal-processing section 106 and a disk RAM 105. A disk controller 104 is equipped with the ECC section 112, in the case of writing, an error detection correction sign (ECC) is generated / added to transfer data, and is read using this ECC in the case of read-out, and performs error detection / correction processing to data to it.

[0023] The signal-processing section 106 performs encoding/decoding, and A/D conversion of data. The HDA (Hard Disk Assembly) section 108 consists of voice coil motors for moving the record medium which stores data, the spindle motor made to rotate a record medium, a lead/write head, the actuator supporting a lead/write head, and an actuator etc. The servo controller 107 performs control of a spindle motor and a voice coil motor.

[0024] Next, the technique of attaining optimization of the information processing system shown in drawing 1 is explained. This technique is setting up the suitable

value for the parameter (engine-performance parameter) which determines the engine performance of a hard disk drive unit 101, and attains optimization of the information processing system containing a hard disk drive unit 101.

[0025] First, the engine-performance parameter of a hard disk drive unit 101 is explained.

[0026] Drawing 2 is drawing showing the example of the engine-performance parameter of a hard disk drive unit 101. Hereafter, each parameter is explained.

[0027] Spindle rotational speed (offset address: 0) shows the rotational speed of the spindle motor which fixes a record medium. In this example, the spindle motor of a hard disk drive unit 101 is supporting three kinds of rotational speed (3600/5400/7200).

[0028] The cache control mode (offset address: 1) shows the control system of the segment managed on a disk RAM 105. In a hard disk drive unit 101, one of the fixed segment method which fixes segment length beforehand, and the adjustable segment methods from which segment length changes with the access patterns from host equipment 113 can be chosen.

[0029] Cache size (offset address: 2) shows the cache size on a disk RAM 105. In this example, four kinds of cache sizes (0.5/1.0/2.0/4.0) are supported, and it

is usually set as 4MB which is the size of the physical disk RAM 105. Compared with the throughput of host equipment 113, the data transfer capacity of a hard disk drive unit 101 is too high, and it is possible to reduce the effective transfer rate of a hard disk drive unit 101 to the case where the system-wide engine performance falls, by making cache size small and reducing a hit ratio intentionally.

[0030] The number of the maximum read-ahead sectors (offset address: 3) shows continuously the number of max sectors predicted on a lead segment, after reading the directed sector data on a lead segment. In this example, four kinds of numbers of max sectors (64/128/256/512) are supported.

[0031] Seeking mode (offset address: 4) shows the seeking method which moves a lead/write head to the target sector. A lead/write head is equipped with the fast mode which moves to a high speed, and the slow mode which moves to a low speed in this example.

[0032] A reserve sector arrangement method (offset address: 5) shows in what kind of unit a reserve sector is arranged. It is possible to set up by any it shall arrange between a track unit, a cylinder unit, a zone unit, and an equipment unit in this example.

[0033] The number of reserve sectors (offset address: 6) is the number of the reserve sectors set up for every reserve sector arrangement unit.

[0034] ECC length (offset address: 7) shows the die length of error detection / correction sign. In this example, four kinds of code length (10/20/30/40) is supported.

[0035] The number of on-the-fly ECC corrections (offset address: 8) shows the correction block count of On-the-fly ECC. In this example, even 1-4 are selectable. On-the-fly ECC corrects error data, continuing consecutive data transfer, even when an error is in read-out data.

[0036] A host transfer rate (offset address: 9) shows the maximum data transfer rate between the host interface controller 117 and the disk interfacing controller 103. In this example, four kinds of transfer rates (16/33/66/100) are supported.

[0037] With this operation gestalt, the user of host equipment 113 can unite and set an engine-performance parameter which was mentioned above as liking of a user. Under the present circumstances, since it is easy, it is made to make a user specify a mode of operation here, although a user may be made to choose a desired value about each engine-performance parameter.

[0038] Drawing 3 is drawing showing the example of a mode of operation. As

shown in this drawing, there are the high performance mode 501, the low noise / low-power mode 502, the animation playback mode 503, and high reliance mode 504 in a mode of operation in this case.

[0039] In the high performance mode 501 (set point "aaa"), in order to think the engine performance as important, the maximum which equipment can take is set as each parameter, such as spindle rotational speed, cache size, seeking mode, and a host transfer rate.

[0040] In the low noise / low-power mode 502 (set point "bbb"), in order to stop the noise sound and power consumption which are generated from a hard disk drive unit 101 as much as possible, a necessary minimum value when performing application is set as each parameter of spindle rotational speed, seeking mode, and a host transfer rate.

[0041] In the animation playback mode 503 (set point "ccc"), in order to think a real-time transfer of read-out data as important, according to the band which host equipment 113 needs, each parameter, such as spindle rotational speed, seeking mode, and a host transfer rate, is set up. Moreover, although not shown in drawing, even when the error in which ECC correction is impossible occurs, it is also effective to form the mode which continues consecutive data transfer.

[0042] In the high reliance mode 504 (set point "ddd"), in order to think the dependability of data as important, the minimum value is set as each parameter, such as spindle rotational speed, seeking mode, and a host transfer rate. Moreover, it is also possible by restricting cache size and the number of on-the-fly ECC corrections to raise reliability.

[0043] Next, the processing which sets up the engine-performance parameter of a hard disk drive unit 101 is explained.

[0044] Drawing 4 is drawing showing the flow chart of the processing which sets up the engine-performance parameter of a hard disk drive unit 101 in initialization of a system performed at the time of the reboot by the power up of host equipment 113, or reset.

[0045] As shown in this drawing, host equipment 113 publishes the Identify Device command (code: ECh) first, in order to read the engine-performance parameter of a hard disk drive unit 101 (S202). This command is a command specified by ATA specification.

[0046] A disk CPU 102 will grasp the contents and a hard disk drive unit 101 will return the engine-performance parameter of self to host equipment 113, if the Identify Device command is received. With this operation gestalt, the information

which shows the existence of an engine-performance optimization function, and each engine-performance parameter information are added to the vendor unique field (or reserve field).

[0047] Drawing 5 is drawing showing the information which a hard disk drive unit 101 returns to host equipment 113 to the Identify Device command. As shown in this drawing, the base address (y) of the information which shows the existence of an engine-performance optimization function, and an engine-performance parameter storing field was added to WORD x, and each engine-performance parameter information is added to it after WORD y.

[0048] Host equipment 113 will distinguish whether the hard disk drive unit 101 is supporting the engine-performance optimization function, if the response to the Identify Device command is received from a hard disk drive unit 101 (S203). Consequently, when not supporting, (S203:No) and a default setup are used (S204). On the other hand, when supporting, information is read in the storing field (Word y-) of (S203:Yes) and each engine-performance parameter, and the range which can be set up and initial value of each engine-performance parameter are saved (S205).

[0049] And when distinguishing whether an above-mentioned

engine-performance parameter is made to set up to the user of host equipment 113 (S206) and making him set it as a user, the mode of operation of a request of a hard disk drive unit 101 is asked to (S206:Y) and a user (S207).

[0050] If a desired mode of operation is chosen by the user (S208), it directs that the Set Feature command is published (S209) and modification of an engine-performance parameter becomes the selected mode of operation, and the parameter after changing into coincidence is saved (S210).

[0051] The Set Feature command was prescribed by ATA specification, is a command and has added the instruction which directs a setup of an engine-performance parameter to a vendor unique field (or reserve field) with this operation gestalt.

[0052] If a hard disk drive unit 101 receives the Set Feature command, a disk CPU 102 will grasp the contents and the existence of the change request to an engine-performance parameter will be checked. When there is a change request, a disk CPU 102 issues modification directions to each internal control block which is performing control of a hard disk drive unit 101 according to the parameter with which modification was demanded. For example, modification to spindle rotational speed performs control which it was notified to the spindle

motor of the servo controller 107 and the HDA section 108, and was doubled with the demanded rotational speed. When it is changed into 7200rpm from 3600rpm in this case, the sampling period of a servo sector is set as one half. This programs the control approach for each set point beforehand, and when this set point is supplied, it should just perform the program over this set point. The cache control mode, cache size, the number of the maximum read-ahead sectors, ECC length, and modification to the number of on-the-fly ECC corrections are notified to a disk controller 104, and control to each desired value is performed. Moreover, seeking mode is notified to the servo controller 107, and a host transfer rate is notified to the disk interfacing controller 103. Moreover, about a reserve sector arrangement method and the number of reserve sectors, it is notified to a disk CPU 102.

[0053] In addition, when making the mode of operation shown in drawing 3 intermingled in one hard disk drive unit, you may make it decide the storing field of the hard disk drive unit corresponding to each mode. For example, a periphery zone will become possible [efficient usage], if it carries out using it for high performance mode or an animation playback mode etc.

[0054] Although the user is performing a setup of an engine-performance

parameter at the time of initialization of a system, you may make it host equipment 113 reset automatically during actual data processing in the example mentioned above according to the contents. For example, what is necessary is just to set it as the animation playback mode 503, when setting it as the low noise / low-power mode 502 when calling and hearing the music data stored in the hard disk drive unit 101 from host equipment 113, or seeing an animation.

[0055] Drawing 6 is drawing showing the flow chart of the processing for which host equipment 113 resets the engine-performance parameter of a hard disk drive unit 101 according to application.

[0056] In order to move the started application if a certain application is started (S602) as shown in this drawing, the optimal mode of operation is chosen (S603). In case application software is installed, matching with each application and a mode of operation may be performed according to the default setting of application, and may be performed according to a user's selection.

[0057] Host equipment 113 changes the value of the engine-performance parameter saved at the time of initialization, like the time of initialization, publishes the Set Feature command and directs modification of an engine-performance parameter so that the selected mode of operation may be

realized (S604). And in order to move application, application is performed using the hard disk drive unit 101 which became the optimal mode of operation (S605). After ending activation of application (S607:Y), again, host equipment 113 publishes the Set Feature command, and returns an engine-performance parameter to initial value (S608).

[0058] As explained above, with this operation gestalt, it becomes possible to attain system optimization containing a hard disk drive unit 101 by setting up the suitable value for the engine-performance parameter which determines the engine performance of a hard disk drive unit 101.

<< -- 2nd operation gestalt>> -- next, the 2nd operation gestalt of this invention is explained.

[0059] Drawing 7 is drawing showing the configuration of the 2nd information processing system which applied this invention. This structure of a system is the same as what was shown in the disk interfacing controller 103 of a hard disk drive unit 101 at drawing 1 except for the point of having formed the host effective transfer rate calculation section 109.

[0060] With this operation gestalt, host equipment 113 recognizes the engine-performance parameter of a hard disk drive unit 101 like the 1st

operation gestalt. And optimization as the whole system is attained by predicting in advance the processing time which command execution takes in host equipment 113 using this parameter, and rearranging the issue sequence of a command based on a prediction result.

[0061] With this operation gestalt, in order to predict the processing time which command execution takes, in addition to the engine-performance parameter shown with the 1st operation gestalt, a seek-time formula and sector-address transformation are notified to host equipment 113 from a hard disk drive unit 101 as an engine-performance parameter.

[0062] A seek-time formula is a parameter for specifying the formula for computing time amount required for a lead/write head moving to the target sector.

[0063] Sector-address transformation is a parameter which specifies the transformation which computes the physical address information (a cylinder, a head, sector address) actually accessed from LBA (logicalblock address) given from host equipment 113. In the transformation concerned, the number of sectors per track for every zone is also taken into consideration.

[0064] Moreover, in the 1st operation gestalt, as one of the engine-performance

parameters, although the host transfer rate was mentioned, this is the maximum data transfer rate on an interface, and in fact, since it is accompanied by preparation of data transfer in host equipment 113 and a hard disk drive unit 101, the data transfer between both is not necessarily transmitted with the maximum data transfer rate.

[0065] So, with this operation gestalt, in order to enable prediction of a more exact data transfer time, the host effective transfer rate calculation section 109 is formed in the disk interfacing controller 103. The host effective transfer rate calculation section 109 consists of a sector counter 111 which counts the transmitted number of sectors, and a timer 110 which measures the time amount concerning the data transfer concerned, and measures an effective transfer rate in each data transfer.

[0066] The measured effective transfer rate may be notified to host equipment 113 at any time, may perform statistics processing with a disk CPU 102, and may notify it to host equipment 113 as a statistical data. In addition, the host effective transfer rate calculation section 109 may be formed in the host equipment 113 side.

[0067] Drawing 8 is drawing showing a flow chart in case host equipment 113

publishes input/output request (ATA command) to a hard disk drive unit 101.

[0068] In the environment where two or more applications are processed by coincidence like multitasking OS, the issue demand of the ATA command is performed for every application.

[0069] First, if the application which the host CPU 114 performs is chosen (S702), in order that the host CPU 114 may perform radial transfer of data required for activation of the application if needed, issue of the ATA command is required from the host interface controller 117 (S703).

[0070] Such a command issue demand is stored in the command queue 801 (S704). About the detail of the command queue 801, it mentions later.

[0071] Next, command execution time amount is predicted about each command collected on the command queue 801 using an above-mentioned engine-performance parameter (S705), and command execution sequence is rearranged so that a system may become the optimal (S706).

[0072] And sequential execution of the rearranged command is carried out (S707). The above processings are repeated until all applications are completed.

[0073] In addition, if possible when a new command is inputted into the command queue 801 into command execution, it will rearrange by judging the

dispatching priority of the command.

[0074] Next, the detail of the command queue 801 mentioned above is explained.

[0075] Drawing 9 is drawing showing the configuration of the command queue 801. As mentioned above, the command issue demand from each application is stored in the command queue 801.

[0076] The command issue demand has the band guarantee flag 802 which shows whether the command needs the band guarantee in addition to the class 803 of command, and the information for [804 (Initiation LBA, the number of access blocks)] access. In case the band guarantee flag 802 transmits animation information, when a data transfer band always needs to be guaranteed, it is set up by application. For example, a band guarantee responds for whether being the need or not, and, as for the band guarantee flag 802, 0 or 1 are set up. In addition, the priority and permission command execution time amount of a band guarantee may be included in addition to this.

[0077] In the command queue 801, the command by which the issue demand was carried out newly is stored in the tail end of the command queue 801, sequential execution is carried out from a forefront command, and the executed command is taken out out of a queue.

[0078] Next, rearrangement (S706) of the command mentioned above is explained to a detail.

[0079] Drawing 10 and drawing 11 are drawings showing the flow chart of the processing which rearranges an above-mentioned command.

[0080] First, it investigates whether a lap is in the range for access in the command stored in the command queue 801 (S902). Consequently, since a just processing result will not be obtained if the execution sequence of a command is replaced when there is a lap (S902:Y), it blocks so that the execution sequence of a command group with a lap may not be replaced (S903).

[0081] Next, it investigates whether there is any light command with a band guarantee (S904). Consequently, when there is the command concerned, (S904:Y) and the command group concerned are put into the head of a queue 801 (S905).

[0082] Next, it investigates whether there is any lead command with a band guarantee (S906). Consequently, when there is the command concerned, (S906:Y) and the command group concerned are put into a continuation of a queue 801 (S907).

[0083] To the command group which remained, the physical cylinder address

which serves as a candidate for access from LBA of a command is computed using the sector-address transformation contained in the above-mentioned engine-performance parameter, command sequence is rearranged so that it may become the shortest, the cylinder migration length, i.e., the seeking distance, between command executions, and it puts into a queue 801 (S908).

[0084] Next, the execution time to a light command group with a band guarantee is predicted using an engine-performance parameter (S909), and it judges whether the guarantee of a band is possible (S910). Consequently, error processing is performed when it cannot guarantee (i.e., when not catching up with the band where the throughput of a hard disk drive unit 101 was demanded) (S910:N) (S917). In error processing, activation of the command concerned is stopped in advance, a user is notified or the change request (slowdown) of an effective transfer rate is performed to information dispatch devices (for example, video camera connected to host equipment 113).

[0085] On the other hand, when a band guarantee is possible, the command from which the disk RAM 105 of a hard disk drive unit 101 will be in a full condition with the transfer data from host equipment 113 is detected from the prediction result of (S910:Y) and above-mentioned command execution time

amount (S911). And if the lead command hit to the data on a disk RAM 105 is in the command queue 801, the lead command concerned will be inserted after the detected command (S912). In this case, if the transfer data stored on the disk RAM 105 are transmitted to the HDA section 108 and an undershirt run does not occur, two or more commands which carry out a lead hit may be inserted.

[0086] Next, similarly, the execution time to a lead command group with a band guarantee is predicted using an engine-performance parameter (S913), and it judges whether the guarantee of a band is possible (S914). Consequently, when it cannot guarantee, error processing of stopping activation of the command concerned is performed like the time of (S914:N) and an above-mentioned light (S917). On the other hand, when it can guarantee, the command to which the disk RAM 105 of a hard disk drive unit 101 will be in a full condition from the prediction result of (S914:Y) and above-mentioned command execution time amount with the transfer data from the HDA section 108 is detected (S915). And if there is a lead command which is a lead command in the command queue 801, and is hit to the data on a disk RAM 105, it will insert after the lead command which had the lead command concerned detected (S916). In this case, if the transfer data stored on the disk RAM 105 are transmitted to host equipment 113

and an undershirt run does not occur, two or more above-mentioned commands which carry out a lead hit may be inserted.

[0087] Rearrangement of a command is performed as mentioned above.

[0088] Next, the concrete forecasting method of above-mentioned command execution time amount is explained. Command execution time amount is computed by dividing it in the following three cases.

(a) Since the hard disk drive unit 101 has adopted the write cache control system in the case of a light command, when requested data is transmitted to a disk RAM 105 from the host interface controller 117, it is regarded as the Commando **. Therefore, if n and the host effective transfer rate at the time of an above-mentioned light are made into X_w and 1 sector = 512 byte for the number of transfer sectors, the command execution time amount T is $T = (512 \times n) / X_w \dots$

(1)

It becomes.

(b) It is a lead command, and when carrying out a cache hit, when requested data is transmitted to the host interface controller 117 from a disk RAM 105, it is regarded as the Commando ** like (a). Therefore, the command execution time amount T is [X_r , then] $T = (512 \times n) / X_r$ about the host effective transfer rate at the

time of a lead... (2)

It becomes.

(c) Since it is necessary to read requested data from a record medium in this case when it is a lead command and makes a cache mistake, the command execution time amount T is [0089].

[Equation 1]

$$T = T_s + T_r + (60/X_s) \times (n/S) + (512 \times 1)/X_r \dots (3)$$

seek-time: -- T_s [second] rotational-delay: -- T_r [second] spindle
rotational-speed: -- X_s [rpm]

The number of sectors per track: It is set to S.

[0090] In addition, since the data transfer by the side of a host becomes a neck when reverse although the case where a host effective transfer rate is quicker than the data read-out rate from a record medium is assumed, a formula (3) is [0091].

[Equation 2]

$$T = T_s + T_r + (60/X_s) \times (1/S) + (512 \times n)/X_r \dots \text{It becomes (3)'}$$

[0092] The amount of seeking to the purpose sector is computed using above-mentioned sector-address transformation, and the seek time T_s can be

asked from a seek-time formula. Moreover, rotational delay may find average rotational delay from spindle rotational speed, or may measure the latency time with a hard disk drive unit 101. In addition, although the case where it accesses ranging over two or more tracks by one command is not assumed by the formula (3), since the time amount which carries out a track change from a seek-time formula even in such a case is computable, prediction of the execution time is possible.

[0093] Drawing 12 is drawing showing the example which replaced the execution sequence of a command based on an above-mentioned command rearrangement flow chart. This drawing (a) shows the situation before rearrangement, and this drawing (b) is drawing showing the situation after rearrangement. In this drawing, RW expresses a random light and, as for SW, RR expresses a random lead for a sequential light. The figure after SW etc. expresses the sequence that each command was published. For example, SW0, SW1, SW2, and SW3 express four sequential lights published in this sequence. Moreover, the hit flag 1001 shows the result of having predicted whether the object data of each command having existed on the disk RAM 105 of a hard disk drive unit 101, in host equipment 113 based on the engine-performance

parameter and the published command issue requested data of a hard disk drive unit 101.

[0094] As shown in drawing 10 (b), as a result of rearranging the command group shown in this drawing (a), priority is first given to a light command with a band guarantee (1). Next, since it was predicted that a disk RAM 105 will be in a full condition when the command SW2 was published as a result of predicting command execution time amount from an engine-performance parameter, commands RR0 and RR2 (2) are inserted after a light command with a band guarantee. It is expected that these commands are hit from an engine-performance parameter to the data on a disk RAM 105. finally, rearrangement **** and the remaining command (3) are placed so that seeking distance may become the shortest.

[0095] As mentioned above, when host equipment 113 predicts command execution time amount from the engine-performance parameter which determines the engine performance of a hard disk drive unit 101 and rearranges the execution sequence of a command with this operation gestalt, it is possible to attain system optimization.

<< -- 3rd operation gestalt>> -- next, the 3rd operation gestalt of this invention is

explained.

[0096] Drawing 13 is drawing showing the home network structure of a system which applied this invention. As shown in this drawing, this system consists of a set top box 1203, a digital television 1210, and a hard disk drive unit 101. Each devices 101, 1203, and 1210 are connected by IEEE1394.

[0097] A set top box 1203 is equipped with a tuner 1204, A/D converter 1205, a demodulator 1206, a descrambler 1207, the 1394 interface sections 1201, and CPU1208 and RAM1209.

[0098] After a set top box 1203 detects the information inputted from the outside, such as cable broadcast and satellite broadcasting service, with a tuner 1204 and changes it into digital information with A/D converter 1205, it is outputted to other 1394 devices through a demodulator 1206, a descrambler 1207, and the IEEE1394 interface section 1201.

[0099] A digital television 1210 is equipped with the 1394 interface sections 1201, the MPEG decoder 1211, the display control section 1212, and the display section 1213.

[0100] A digital television 1210 decodes the information inputted through the 1394 interface sections 1201 by the MPEG decoder 1211, and displays it in the

display section 1213 through the display control section 1212.

[0101] Although a hard disk drive unit 101 does not change as fundamentally as what was shown in drawing 1 and drawing 7 , it is equipped with the 1394 interface sections 1201 instead of the disk interfacing controller 103 in order to control an IEEE1394 interface.

[0102] With this operation gestalt, the self ID packet transmitted and received at the time of initialization of 1394 interfaces is used as a way method which notifies the engine-performance parameter of 1394 devices connected to the network of hard disk drive unit 101 grade to other 1394 devices.

[0103] Drawing 14 is drawing showing the flow chart of initialization in 1394 interfaces.

[0104] If bus reset occurs in connection with the power-on/OFF of 1394 devices, or plug-in/out, initialization of a bus will be performed and all topology information will be eliminated (S1302). Next, in order to acquire new topology information, a tree ID process is performed (S1303). Here, the device used as the parentage between devices and the root is determined. The device information which transmits and receives a self ID packet and finally contains the physics ID of a device is exchanged with each other (S1304).

[0105] Drawing 15 is drawing showing the configuration of a self ID packet. With this operation gestalt, in order to notify the engine-performance parameter of a hard disk drive unit 101 to other 1394 devices, the packet of the packet number n is made into the packet for notifying engine-performance parameter information. And if the contents of the bit 0 of this packet are "1", it will be specified that the device has the function which notifies an engine-performance parameter. Moreover, IEEE1212 by which the engine-performance parameter is stored in bits 1-31 The address offset value 1401 in the address space based on CSR (control & status register) architecture is put in. In addition, you may make it put engine-performance parameter value into a bit 1 - 31 grades.

[0106] In the IEEE1394 interface, the 64-bit address space specified in IEEE1212 is supported. Drawing 16 is drawing showing the 64-bit address space specified in IEEE1212. By IEEE1212 specification, Bus ID is specified in 10 bits of high orders of the 64-bit address, Node ID is specified by the next 6 bits, and the address in each node is specified by the remaining 48 bits.

[0107] It arranges to the initial unit tooth space, and a hard disk drive unit 101 puts the offset 1401 of the arranged address into the bits 1-31 of the above-mentioned self ID packet number n for the engine-performance

parameter information on self. The device which received the self ID packet which the hard disk drive unit 101 transmitted transmits a lead block request packet to a hard disk drive unit 101, and reads the engine-performance parameter stored in the offset address 1401.

[0108] In the system shown in drawing 13 , a set top box 1203 and a digital television 1210 receive the self ID packet transmitted from the hard disk drive unit 101. By receiving this self ID packet, these devices can grasp that the hard disk drive unit 101 has the function which notifies an engine-performance parameter, and can also receive an engine-performance parameter further.

[0109] Next, rearrangement of the command in this operation gestalt is explained.

[0110] With this operation gestalt, since two or more devices which access a hard disk drive unit 101 exist, rearrangement of a command chooses a master device from two or more devices which publish a command to a hard disk drive unit 101, and is performed in a master device. With this operation gestalt, a set top box 1203 presupposes that the role of a master device is taken charge of. In addition, if it is the device connected with the network, the same control is possible from properties (peer to peer), such as a device pair which 1394

interfaces have.

[0111] 1394 interfaces have the function which chooses the isochronous manager who manages an isochronous transfer in the case of initialization of a bus. An isochronous manager performs management for guaranteeing the transfer band of the packet transmission which needs the time amount guarantee of an animation etc. After a device to transmit notifies the bandwidth needed for an isochronous manager and obtains authorization, it transmits an isochronous packet. With this operation gestalt, the device chosen as the isochronous packet manager has the role of a master device.

[0112] Devices other than the master device which wants to access a hard disk drive unit 101 notify that to a master device, and transmit a command group required for the processing. What is necessary is just to perform this using the usual asynchronous packet.

[0113] For example, when a digital television 1210 wants to access a hard disk drive unit 101, a digital television 1201 will access a hard disk drive unit 101 via a set top box 1203. First, a digital television 1201 sends a packet to a certain specific address space to a set top box 1203. This specific address space notifies using it for the packet issue request which went via the master device

beforehand to all 1394 devices. And the issue client (digital television 1201 in this case) of a packet, the packet destination (hard disk drive unit 101 in this case), and the contents of the command are stored in the data division of this packet. The set top box 1203 which is the master device which received this packet transmits the demand from a packet issue client to the packet destination. About the check processing in transmission and reception of a packet, as 1394 interfaces have prescribed, to perform Acknowledgement processing at the time of an asynchronous transfer, and what is necessary is made just not to perform check processing at the time of an isochronous transfer.

[0114] A set top box 1203 can adjust the command group and issue sequence which were received from other devices in response to the transmitted command group, and can attain optimization of system performance.

[0115] Drawing 17 is drawing showing the example of rearrangement of a command like drawing 12 . Here, the case where the writing and read-out processing of random data to a hard disk drive unit 101 are performed to coincidence is assumed to be the processing which reads an animation from a hard disk drive unit 101, and is outputted to a digital television 1210 from the set top 1203.

[0116] As shown in drawing 17 , the sequential lead command (SRx) with a band guarantee which a digital television 1201 publishes to a hard disk drive unit 101, and the demand of the random access (RRx, RWx) from a set top box 1203 are coming.

[0117] Although the head of the command queue 801 before rearrangement is a random lead, in order to already ** the animation playback response to a digital television 1210, top priority is given to SR0, and a read ahead (data transfer from the HDA section 108 to a disk RAM 105) is performed (1). Since SR0 does not have object data and serves as read-out from the HDA section 108 on a disk RAM 105, it requires time amount for seeking etc. Therefore, RR0 and RR1 are performed next (2). Since this is carrying out the cache hit and becomes the data transfer to a set top box 1203 from a disk RAM 105, said read ahead is continued. Next, SRs 1-3 are performed (3). Since the data for a transfer are already stored in the disk RAM 105, SR1 serves as a lead hit. Moreover, since RAM of the 1394 interface sections 1201 by the side of a digital television 1210 will be in a full condition by the data transfer by SR2 and SR3, it can reproduce, without an animation breaking off, even if it performs RW0 and RR2 (4) before SR4.

[0118] Although the execution time of the command published by the set top box 1203 at a hard disk drive unit 101 based on delivery and the engine-performance parameter of a hard disk drive unit 101 in the engine-performance parameter of a hard disk drive unit 101 was predicted above, you may make it notify the engine-performance parameter of a digital television 1210 to a set top box 1203 further.

[0119] In this case, a set top box 1203 becomes possible [predicting more the execution time of the command published to a hard disk drive unit 101 to accuracy]. For example, if the capacity of RAM1202 of the 1394 interface sections 1201 of a digital television 1210, the data output speed from RAM1202 to the MPEG decoder 1211, and the data input speed to RAM1202 are notified, in a set top box 1203, it can predict when RAM1202 becomes full by the data transfer to RAM1202.

[0120] When the device which accesses a hard disk drive unit 101 with this operation gestalt is plurality, and a master device is chosen from two or more devices which publish a command to a hard disk drive unit 101, and the master device concerned carries out package management of the access to a hard disk drive unit 101, predicts command execution time amount and rearranges the

execution sequence of a command, it becomes [as explained above,] possible to attain system optimization.

[0121] Moreover, also in the packet transfer which needs the band guarantee of an animation transfer etc., it is the form where system states, such as an internal state of not only the band guarantee on an IEEE1394 bus but a hard disk drive unit, were taken into consideration, and it is possible for the band of data transfer to be guaranteed and to attain system optimization.

[0122] In addition, the hard disk drive unit 101 shown with the 2nd and 3rd operation gestalt may use what has an engine-performance parameter strange good like the 1st operation gestalt. In this case, host equipment etc. saves the engine-performance parameter value after modification, and should just be made to perform prediction of command execution time amount etc. using the engine-performance parameter value after modification.

[0123]

[Effect of the Invention] As explained to the detail above, it is possible to attain system optimization containing external storage by setting up the suitable value for the engine-performance parameter which determines the engine performance of external storage in this invention.

[0124] Moreover, it becomes possible by predicting command execution time amount from engine-performance parameters, such as external storage, and rearranging the execution sequence of a command to aim at improvement in system performance.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] It is the block diagram of the information processing system by this invention.

[Drawing 2] It is drawing showing the example of the engine-performance parameter of a hard disk drive unit.

[Drawing 3] It is drawing showing the example of a mode of operation.

[Drawing 4] It is the flow chart which shows setting processing of an engine-performance parameter.

[Drawing 5] It is drawing showing the configuration of Identify Device information.

[Drawing 6] It is the flow chart which shows setting processing of the engine-performance parameter at the time of application activation.

[Drawing 7] It is the block diagram of the 2nd information processing system by this invention.

[Drawing 8] It is the flow chart which shows command execution processing.

[Drawing 9] It is drawing showing the configuration of a command queue.

[Drawing 10] It is the flow chart (the 1) which shows command rearrangement

processing.

[Drawing 11] It is the flow chart (the 2) which shows command rearrangement processing.

[Drawing 12] It is drawing showing an example of command rearrangement.

[Drawing 13] It is the block diagram of the home network system by this invention.

[Drawing 14] It is drawing showing the flow chart of initialization processing of an IEEE1394 interface.

[Drawing 15] It is drawing showing the example of a format of a self ID packet.

[Drawing 16] It is the conceptual diagram of the address space used with an IEEE1394 interface.

[Drawing 17] It is drawing showing an example of command rearrangement.

[Description of Notations]

101 [-- A command queue, 802 / -- A band guarantee flag, 1201 / -- The 1394 interface sections, 1203 / -- A set top box, 1210 / -- Digital television] -- A hard disk drive unit, 103 -- A disk interfacing controller, 113 -- Host equipment, 801

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2001-222380
(P2001-222380A)

(43) 公開日 平成13年8月17日 (2001.8.17)

(51) Int.Cl. ⁷	識別記号	F I	テ-マコ-ト (参考)
G 0 6 F 3/06	3 0 1	G 0 6 F 3/06	3 0 1 M 5 B 0 6 5
G 1 1 B 19/02	5 0 1	G 1 1 B 19/02	5 0 1 K 5 D 0 4 4
20/10		20/10	D

審査請求 未請求 請求項の数 5 O L (全 16 頁)

(21) 出願番号 特願2000-29286 (P2000-29286)

(22) 出願日 平成12年2月7日 (2000.2.7)

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 角田 元泰

神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

(72) 発明者 平塚 幸恵

神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

(74) 代理人 100087170

弁理士 富田 和子

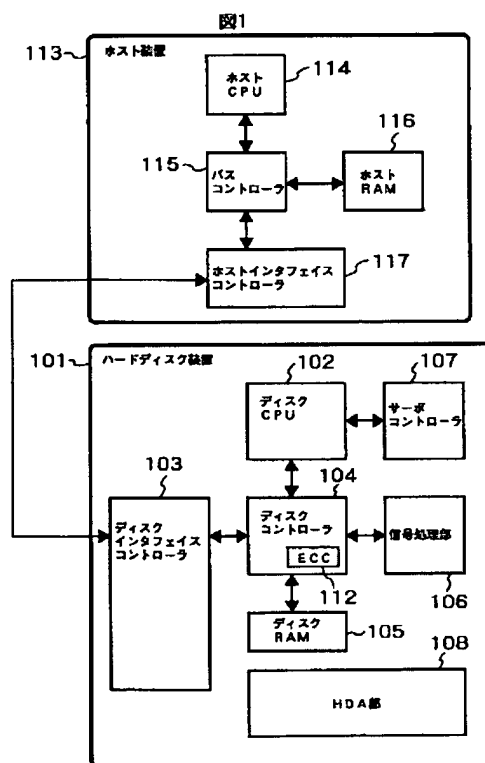
最終頁に続く

(54) 【発明の名称】 外部記憶装置とそれを備えた情報処理システム

(57) 【要約】

【課題】 ハードディスク装置等の外部記憶装置を含む情報処理システムの最適化を図る。

【解決手段】 ハードディスク装置101とホスト装置113との間で、ハードディスクの性能を決定付ける性能パラメータ（スピンドル回転速度、キャッシュサイズ等）を送受信する。ホスト装置113から、性能パラメータの設定要求を受けると、ハードディスク装置101は、指示された性能パラメータの再設定を行う。また、ホスト装置113は、ハードディスク装置101から受信したパラメータ情報に基づいて、ハードディスク装置101に対するコマンドの実行時間を予測し、この予測結果に基づいて、コマンドの並び替えを行う。



【特許請求の範囲】

【請求項1】 情報処理装置がアクセスを行う外部記憶装置であって、

前記情報処理装置との間で、自己の性能を決定付ける性能情報を送受信する手段と、

前記情報処理装置から、前記性能情報の値の変更を要求を受けて、自己の性能情報の値の変更を行う手段とを備えたことを特徴とする外部記憶装置。

【請求項2】 請求項1に記載の外部記憶装置と、当該外部記憶装置に対してアクセスを行う情報処理装置とを備えたことを特徴とする情報処理システム。

【請求項3】 外部記憶装置にアクセスを行う情報処理装置であって、

前記外部記憶装置から、当該外部記憶装置の性能を決定付ける性能情報を受信する手段と、

前記外部記憶装置に対するコマンドを複数個蓄えておくコマンドキュー手段と、

前記外部記憶装置から受信した性能情報に基づいて、前記コマンドキュー手段に蓄えられた各々のコマンドについてコマンド実行時間を予測する手段と、

予測されたコマンド実行時間に基づいて、システムが最適化されるようにコマンドキュー手段に格納されたコマンドの実行順序を並べ替える手段とを備えたことを特徴とする情報処理装置。

【請求項4】 前記外部記憶装置にアクセスする他の情報処理装置から、前記外部記憶装置に対するコマンドを受信し、前記コマンドキュー手段に格納する手段を更に備えることを特徴とする請求項3に記載の情報処理装置。

【請求項5】 請求項3又は請求項4に記載の情報処理装置と、

前記性能情報を前記情報処理装置に送信する外部記憶装置とを備えたことを特徴とする情報処理システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、ハードディスク装置等の外部記憶装置を含む情報処理システムの最適化に関する。

【0002】

【従来の技術】近年、ハードディスク装置の大容量化、高速化に伴い、動画や音声をハードディスク装置に記録／再生するシステムが提案されている。動画や音声といったデータは、リアルタイム処理、即ち、データ転送の帯域保証が必要になる。

【0003】そのため、例えば、特開平10-222310号公報に記載されている磁気ディスク装置では、実行時間予測を行う入出力コマンドをホストコンピュータから受けると、同コマンド処理に必要な実行時間を予測し、その結果がホストコンピュータの要求している最大許容時間内であればコマンドを実行し、許容時間をオー

バーするようであればコマンドの実行を中止して、ホストコンピュータにその旨を通知していた。

【0004】

【発明が解決しようとする課題】上記公報記載の技術は、データ転送の帯域保証に主眼を置いており、ディスク装置それ自体は内部状態からコマンド実行時間を予測するだけである。従って、ディスク装置がシステムの中で最適な状態で稼働しているとは限らない。例えば、音声を再生する場合、仮にディスク装置の転送帯域が十分に保証されているならば、スピンドル回転速度を落としてディスク装置の発するノイズ音を低減することが望ましい場合もある。

【0005】また、上記公報記載の技術では、入出力コマンドが許容時間内に実行できない場合は、その旨がホストコンピュータに通知されるが、ホストコンピュータは、コマンドを実際に発行しないとその結果が得られない。そのため、ディスク装置に対して無駄なコマンドが発行される場合が生じる。また、システム全体でみた場合、ディスク装置に対して、最適なタイミングでコマンドが発行されているとは限らない。

【0006】本発明の目的は、ハードディスク装置等の外部記憶装置を含む情報処理システムが、ユーザの好みに合わせて、若しくは各々のアプリケーションの実行が最適化されるように稼動するための技術を提供することにある。

【0007】また、本発明の他の目的は、システムのパフォーマンスを向上させるために、外部記憶装置に対するコマンドの発行を最適化する技術を提供することにある。

【0008】

【課題を解決するための手段】本発明に係る外部記憶装置は、情報処理装置がアクセスを行う外部記憶装置である。そして、前記情報処理装置との間で、自己の性能を決定付ける性能情報を送受信する手段と、前記情報処理装置から、前記性能情報の値の変更を要求を受けて、自己の性能情報の値の変更を行う手段とを備えたことを特徴とする。

【0009】また、本発明に係る情報処理システムは、前記外部記憶装置と、当該外部記憶装置に対してアクセスを行う情報処理装置とを備えたことを特徴とする。この場合において、情報処理装置は、例えば、システムの初期化時や、アプリケーション実行時に、性能情報の変更を要求する。

【0010】また、本発明に係る情報処理装置は、外部記憶装置にアクセスを行う情報処理装置である。そして、前記外部記憶装置から、当該外部記憶装置の性能を決定付ける性能情報を受信する手段と、前記外部記憶装置に対するコマンドを複数個蓄えておくコマンドキュー手段と、前記外部記憶装置から受信した性能情報に基づいて、前記コマンドキュー手段に蓄えられた各々のコマ

ンドについてコマンド実行時間を予測する手段と、予測されたコマンド実行時間に基づいて、システムが最適化されるようにコマンドキュー手段に格納されたコマンドの実行順序を並べ替える手段とを備えたことを特徴とする。

【0011】この場合において、前記外部記憶装置にアクセスする他の情報処理装置から、前記外部記憶装置に対するコマンドを受信し、前記コマンドキュー手段に格納する手段を更に備えるようにしてもよい。

【0012】また、前記外部記憶装置に対するコマンドは、データ転送の帯域保証が必要であることを示すフラグ情報を有し、当該フラグ情報を用いて、コマンドの実行順序を並び替えるようにしてもよい。更に、コマンド実行時間を予測した結果、前記フラグ情報を有したコマンドについて、データ転送の帯域保証ができない場合は、当該コマンドの実行を中止して、その旨をユーザに通知するようにしてもよい。

【0013】また、前記外部記憶装置又は情報処理装置に、情報処理装置が外部記憶装置に対して発行するコマンドによって生じるデータ転送の量と、データ転送時間から実効データ転送速度を計測する手段を備え、計測結果を使って、コマンドの実行時間を予測するようにしてもよい。

【0014】また、本発明に係る情報処理システムは、前記情報処理装置と、前記性能情報を前記情報処理装置に送信する外部記憶装置とを備えることを特徴とする。

【0015】なお、外部記憶装置がディスク装置の場合、前記性能情報としては、例えば、スピンドル回転速度、キャッシュ制御モード、キャッシュサイズ、最大先読みセクタ数、シーク時間算出式、セクタアドレス変換式、予備セクタ数、ECC長、オンザフライECC訂正数、ホスト転送速度等が挙げられる。

【0016】

【発明の実施の形態】以下、本発明の実施の形態について、図面を参照しつつ、詳細に説明する。

《第1実施形態》まず、本発明の第1の実施形態について説明する。

【0017】図1は、本発明を適用した情報処理システムの構成を示す図である。同図に示すように、本情報処理システムは、ホスト装置113（例えば、パーソナルコンピュータ）と、外部記憶装置としてのハードディスク装置101とから成る。

【0018】本実施形態では、ホスト装置113とハードディスク装置101との間のインタフェースとして、ATA（AT attachment）を想定しているが、他のインタフェース（例えば、IEEE1394や、SSA（serial storage architecture）、Fiber Channel等）も適応可能である。

【0019】ホスト装置113は、ホストCPU114と、バスコントローラ115と、ホストRAM116

と、ホストインタフェースコントローラ117とを備える。なお、当然のことながら、ホスト装置113は、それ以外のコンポーネント、例えば、グラフィックスコントローラや、ネットワークコントローラや、フロッピーディスクコントローラ等を備えるようにしてもよい。

【0020】ホストCPU114は、ROM（不図示）やホストRAM116に格納されたプログラムを実行することにより、所定の処理を行う。バスコントローラ115は、ホストCPU114、ホストRAM116、ホストインタフェースコントローラ117間のデータ転送を制御する。ホストインタフェースコントローラ117は、ハードディスク装置101とホストRAM116との間のデータ転送等を制御する。

【0021】一方、ハードディスク装置101は、ディスクCPU102と、ディスクインタフェースコントローラ103と、ディスクコントローラ104と、ディスクRAM105と、信号処理部106と、サーボコントローラ107と、HDA部108とを備える。

【0022】ディスクCPU102は、ハードディスク装置101全体のデータ処理を制御する。ディスクコントローラ104は、ディスクインタフェースコントローラ103とディスクRAM105との間、及び、信号処理部106とディスクRAM105との間のデータ転送を制御する。ディスクコントローラ104は、ECC部112を備え、書き込みの際には、転送データに対して誤り検出訂正符号（ECC）を生成／付加し、読み出しの際には、このECCを用いて読み出しデータに対する誤り検出／訂正処理を行う。

【0023】信号処理部106は、データのエンコード／デコード処理やA/D変換を行う。HDA（Hard Disk Assembly）部108は、データを格納する記録媒体と、記録媒体を回転させるスピンドルモータと、リード／ライトヘッドと、リード／ライトヘッドを支えるアクチュエータと、アクチュエータを動かすためのボイスコイルモータ等から構成される。サーボコントローラ107は、スピンドルモータやボイスコイルモータの制御を行う。

【0024】次に、図1に示した情報処理システムの最適化を図る手法について説明する。本手法は、ハードディスク装置101の性能を決定付けるパラメータ（性能パラメータ）に適切な値を設定することで、ハードディスク装置101を含む情報処理システムの最適化を図るものである。

【0025】まず、ハードディスク装置101の性能パラメータについて説明する。

【0026】図2は、ハードディスク装置101の性能パラメータの例を示す図である。以下、各パラメータについて説明する。

【0027】スピンドル回転速度（オフセットアドレス：0）は、記録媒体を固定するスピンドルモータの回

転速度を示す。この例では、ハードディスク装置101のスピンドルモータは、3種類の回転速度(3600/5400/7200)をサポートしている。

【0028】キャッシュ制御モード(オフセットアドレス:1)は、ディスクRAM105上で管理されるセグメントの制御方式を示す。ハードディスク装置101では、セグメント長を予め固定する固定セグメント方式と、ホスト装置113からのアクセスパターンによってセグメント長が変化する可変セグメント方式のどちらかを選択できる。

【0029】キャッシュサイズ(オフセットアドレス:2)は、ディスクRAM105上のキャッシュサイズを示す。この例では、4種類のキャッシュサイズ(0.5/1.0/2.0/4.0)をサポートしており、通常は、物理的なディスクRAM105のサイズである4MBに設定する。ホスト装置113の処理能力に比べて、ハードディスク装置101のデータ転送能力が高過ぎて、システム全体の性能が落ちる場合等では、キャッシュサイズを小さくしてヒット率を故意に低下させることによりハードディスク装置101の実効転送速度を落とすことが可能である。

【0030】最大先読みセクタ数(オフセットアドレス:3)は、指示されたセクタデータをリードセグメント上に読み出した後に続けて、リードセグメント上に先読みする最大セクタ数を示す。この例では、4種類の最大セクタ数(64/128/256/512)をサポートしている。

【0031】シークモード(オフセットアドレス:4)は、目的のセクタにリード/ライトヘッドを移動させるシーク方式を示す。この例では、リード/ライトヘッドが高速に移動する高速モードと低速に移動する低速モードを備えている。

【0032】予備セクタ配置方式(オフセットアドレス:5)は、予備セクタを、どのような単位で配置するかを示す。この例では、トラック単位、シリンダ単位、ゾーン単位、装置単位のいずれで配置するかを設定することが可能である。

【0033】予備セクタ数(オフセットアドレス:6)は、各予備セクタ配置単位毎に設定される予備セクタの数である。

【0034】ECC長(オフセットアドレス:7)は、誤り検出/訂正符号の長さを示している。この例では、4種類の符号長(10/20/30/40)をサポートしている。

【0035】オンザフライECC訂正数(オフセットアドレス:8)は、オンザフライECCの訂正ブロック数を示す。この例では、1~4までが選択可能である。オンザフライECCとは、読み出しデータに誤りがある場合でも後続のデータ転送を継続しながら誤りデータの訂正を行うものである。

【0036】ホスト転送速度(オフセットアドレス:9)は、ホストインタフェースコントローラ117とデ

ィスクインタフェースコントローラ103との間の最大データ転送速度を示す。この例では、4種類の転送速度(16/33/66/100)をサポートしている。

【0037】本実施形態では、ホスト装置113のユーザが、上述したような性能パラメータをユーザの好みにあわせて設定することができる。この際、各性能パラメータについてユーザに所望の値を選択させてもよいが、ここでは、簡単のため、ユーザには動作モードを指定させるようにする。

10 【0038】図3は、動作モードの例を示す図である。同図に示すように、この場合、動作モードには、ハイパフォーマンスモード501、低騒音/低消費電力モード502、動画再生モード503、高信頼モード504がある。

【0039】ハイパフォーマンスモード501(設定値「aaa」)では、性能を重視するため、スピンドル回転速度、キャッシュサイズ、シークモード、ホスト転送速度等の各パラメータに、装置の取り得る最大値を設定する。

20 【0040】低騒音/低消費電力モード502(設定値「bbb」)では、ハードディスク装置101から発生するノイズ音や電力消費量を極力抑えるため、スピンドル回転速度、シークモード、ホスト転送速度の各パラメータに、アプリケーションを実行する上で必要最小限の値を設定する。

30 【0041】動画再生モード503(設定値「ccc」)では、読み出しデータのリアルタイム転送を重視するため、ホスト装置113が必要とする帯域に合わせて、スピンドル回転速度、シークモード、ホスト転送速度等の各パラメータを設定する。また図には示していないがECC訂正不可能なエラーが発生した場合でも後続のデータ転送を継続するモードを設けることも有効である。

【0042】高信頼モード504(設定値「ddd」)では、データの信頼性を重視するため、スピンドル回転速度、シークモード、ホスト転送速度等の各パラメータに最小値を設定する。またキャッシュサイズやオンザフライECC訂正数を制限することにより信頼度を上げることも可能である。

40 【0043】次に、ハードディスク装置101の性能パラメータを設定する処理について説明する。

【0044】図4は、ホスト装置113の電源投入時やリセットによる再起動時に行うシステムの初期化において、ハードディスク装置101の性能パラメータを設定する処理のフローチャートを示す図である。

【0045】同図に示すように、ホスト装置113は、まず、ハードディスク装置101の性能パラメータを読み出すために、Identify Deviceコマンド(コード:ECh)を発行する(S202)。同コマンドは、ATA規格で規定されているコマンドである。

【0046】ハードディスク装置101は、Identify Deviceコマンドを受け取ると、ディスクCPU102がその内容を把握し、自己の性能パラメータをホスト装置113に返す。本実施形態では、ベンダユニーク領域（若しくはリザーブ領域）に性能最適化機能の有無を示す情報と、各性能パラメータ情報を付加している。

【0047】図5は、Identify Deviceコマンドに対してハードディスク装置101がホスト装置113に返す情報を示す図である。同図に示すように、ワードxに、性能最適化機能の有無を示す情報及び性能パラメータ格納領域のベースアドレス（y）を付加し、ワードy以降に、各性能パラメータ情報を付加している。

【0048】ホスト装置113は、ハードディスク装置101からIdentify Deviceコマンドに対する応答を受信すると、ハードディスク装置101が性能最適化機能をサポートしているか否かを判別する（S203）。その結果、サポートしていない場合は（S203：No）、デフォルトの設定を使う（S204）。一方、サポートしている場合は（S203：Yes）、各性能パラメータの格納領域（Word y〜）から情報を読み取って、各性能パラメータの設定可能範囲と初期値を保存する（S205）。

【0049】そして、ホスト装置113のユーザに、上述の性能パラメータを設定させるか否かを判別し（S206）、ユーザに設定させる場合は（S206：Y）、ユーザにハードディスク装置101の所望の動作モードを問い合わせる（S207）。

【0050】ユーザによって所望の動作モードが選択されると（S208）、Set Featureコマンドを発行して（S209）、選択された動作モードになるように性能パラメータの変更を指示し、同時に変更後のパラメータを保存しておく（S210）。

【0051】Set Featureコマンドは、ATA規格で規定されてコマンドであり、本実施形態では、ベンダユニーク領域（若しくはリザーブ領域）に性能パラメータの設定を指示する命令を付け加えている。

【0052】ハードディスク装置101は、Set Featureコマンドを受け取ると、ディスクCPU102がその内容を把握し、性能パラメータに対する変更要求の有無を確認する。変更要求があった場合、ディスクCPU102は、変更が要求されたパラメータに応じて、ハードディスク装置101の制御を実行している各々の内部制御ブロックに対して変更指示を出す。例えば、スピンドル回転速度に対する変更は、サーボコントローラ107及びHDA部108のスピンドルモータに通知され、要求された回転速度に合わせた制御を行う。この場合、例えば、3600rpmから7200rpmに変更された場合、サーボセクタのサンプリング周期を半分に設定する。これは、予め各設定値に対する制御方法をプログラミングしておき、この設定値が供給されたときに、この

設定値に対するプログラムを実行すればよい。キャッシュ制御モード、キャッシュサイズ、最大先読みセクタ数、ECC長、及び、オンザフライECC訂正数に対する変更は、ディスクコントローラ104に通知され、各要求値に対する制御が行われる。また、シークモードは、サーボコントローラ107に通知され、ホスト転送速度は、ディスクインタフェースコントローラ103に通知される。また、予備セクタ配置方式及び予備セクタ数については、ディスクCPU102に通知される。

【0053】なお、図3に示した動作モードを1つのハードディスク装置に混在させる場合、各モードに対応したハードディスク装置の格納領域を決めるようにしてもよい。例えば、外周ゾーンは、ハイパフォーマンスモードや動画再生モードに使用する等すれば、効率的な使い方が可能となる。

【0054】上述した例では、性能パラメータの設定は、システムの初期化時にユーザが行っているが、実際のデータ処理中に、その内容に合わせてホスト装置113が自動的に再設定するようにしてもよい。例えば、ハードディスク装置101に格納してある音楽データをホスト装置113から呼び出して聴く場合は、低騒音/低消費電力モード502に設定したり、動画を見る場合は、動画再生モード503に設定すれば良い。

【0055】図6は、ホスト装置113がアプリケーションに合わせてハードディスク装置101の性能パラメータを再設定する処理のフローチャートを示す図である。

【0056】同図に示すように、あるアプリケーションが起動されると（S602）、起動されたアプリケーションを動かすために最適な動作モードが選択される（S603）。各アプリケーションと動作モードとの対応付けは、例えば、アプリケーションソフトをインストールする際に、アプリケーションのデフォルト設定に従って行ってもよいし、ユーザの選択に従って行ってもよい。

【0057】ホスト装置113は選択された動作モードを実現するように、初期化時に保存しておいた性能パラメータの値を変更し、初期化時と同様に、Set Featureコマンドを発行して、性能パラメータの変更を指示する（S604）。そして、アプリケーションを動かすために最適な動作モードになったハードディスク装置101を使って、アプリケーションを実行する（S605）。アプリケーションの実行を終了すると（S607：Y）、ホスト装置113は、再度、Set Featureコマンドを発行して、性能パラメータを初期値に戻す（S608）。

【0058】以上説明したように、本実施形態では、ハードディスク装置101の性能を決定付ける性能パラメータに適切な値を設定することにより、ハードディスク装置101を含むシステムの最適化を図ることが可能になる。

《第2実施形態》次に、本発明の第2の実施形態について説明する。

【0059】図7は、本発明を適用した第2の情報処理システムの構成を示す図である。本システムの構成は、ハードディスク装置101のディスクインタフェースコントローラ103にホスト実効転送速度算出部109を設けた点を除き、図1に示したものと同一である。

【0060】本実施形態では、第1の実施形態と同様に、ホスト装置113がハードディスク装置101の性能パラメータを認識する。そして、同パラメータを使って、ホスト装置113において、コマンド実行に要する処理時間を事前に予測し、予測結果に基づいて、コマンドの発行順序を並び替えることにより、システム全体としての最適化を図るものである。

【0061】本実施形態では、コマンド実行に要する処理時間を予測するため、第1の実施形態で示した性能パラメータに加えて、性能パラメータとして、シーク時間算出式及びセクタアドレス変換式を、ハードディスク装置101からホスト装置113に通知する。

【0062】シーク時間算出式は、目的のセクタにリード／ライトヘッドが移動するのに必要な時間を算出するための算出式を特定するためのパラメータである。

【0063】セクタアドレス変換式は、ホスト装置113から与えられるLBA (logical block address) から実際にアクセスされる物理アドレス情報 (シリンダ、ヘッド、セクタアドレス) を算出する変換式を特定するパラメータである。当該変換式では、各ゾーンごとのトラック当りのセクタ数も考慮されている。

【0064】また、第1の実施形態において、性能パラメータの一つとして、ホスト転送速度を挙げたが、これは、インタフェース上の最大データ転送速度であり、実際には、ホスト装置113及びハードディスク装置101においてデータ転送の準備を伴うため、両者間のデータ転送は必ずしも最大データ転送速度で転送されるわけではない。

【0065】そこで、本実施形態では、より正確なデータ転送時間の予測を可能にするためディスクインタフェースコントローラ103に、ホスト実効転送速度算出部109を設けている。ホスト実効転送速度算出部109は、転送したセクタ数をカウントするセクタカウンタ111と、当該データ転送にかかった時間を計測するタイマ110から構成され、各々のデータ転送において実効転送速度を計測するものである。

【0066】計測された実効転送速度は、随時ホスト装置113に通知しても良いし、ディスクCPU102により統計処理を施し、統計データとしてホスト装置113に通知しても良い。なお、ホスト実効転送速度算出部109を、ホスト装置113側に設けてもよい。

【0067】図8は、ホスト装置113がハードディスク装置101に対して入出力要求 (ATAコマンド) を

発行する場合のフローチャートを示す図である。

【0068】マルチタスクOSのように複数のアプリケーションが同時に処理される環境においては、ATAコマンドの発行要求は、各アプリケーションごとに行われる。

【0069】まず、ホストCPU114が実行するアプリケーションが選択されると (S702)、ホストCPU114は、必要に応じて、そのアプリケーションの実行に必要なデータの入出力処理を行うため、ホストインタフェースコントローラ117に対してATAコマンドの発行を要求する (S703)。

【0070】このようなコマンド発行要求は、コマンドキュー801に格納される (S704)。コマンドキュー801の詳細については、後述する。

【0071】次に、コマンドキュー801に溜まっている各コマンドについて、上述の性能パラメータを用いてコマンド実行時間を予測し (S705)、システムが最適になるようにコマンド実行順序の並べ替えを行う (S706)。

【0072】そして、並べ替えられたコマンドを、順次実行する (S707)。以上のような処理を、すべてのアプリケーションが終了するまで繰り返す。

【0073】なお、コマンド実行中においてもコマンドキュー801に新たなコマンドが入力された場合には、可能であれば、そのコマンドの実行優先順位を判定して並べ替えを行う。

【0074】次に、前述したコマンドキュー801の詳細について説明する。

【0075】図9は、コマンドキュー801の構成を示す図である。前述したように、コマンドキュー801には、各アプリケーションからのコマンド発行要求が格納される。

【0076】コマンド発行要求は、コマンドの種類803、アクセス対象804 (開始LBA、アクセスブロック数) の情報以外に、そのコマンドが帯域保証を必要としているかどうかを示す帯域保証フラグ802を持っている。帯域保証フラグ802は、動画情報を転送する際など、常にデータの転送帯域を保証する必要がある場合にアプリケーションによって設定される。帯域保証フラグ802は、例えば、帯域保証が必要か否かに応じて、0または1が設定される。なお、これ以外に、帯域保証の優先度や、許容コマンド実行時間を含めても良い。

【0077】コマンドキュー801においては、新規に発行要求されたコマンドは、コマンドキュー801の最後尾に格納され、最前のコマンドから順次実行され、実行されたコマンドはキューの外に出される。

【0078】次に、前述したコマンドの並べ替え (S706) について詳細に説明する。

【0079】図10及び図11は、上述のコマンドの並べ替えを行う処理のフローチャートを示す図である。

【0080】まず、コマンドキュー801に格納されたコマンドの中でアクセス対象範囲に重なりがあるか否かを調査する(S902)。その結果、重なりがある場合(S902:Y)、コマンドの実行順序を入れ替えると正当な処理結果が得られないので、重なりがあるコマンド群の実行順序を入れ替えないようにブロックしておく(S903)。

【0081】次に、帯域保証付きのライトコマンドがあるかどうかを調査する(S904)。その結果、当該コマンドがある場合は(S904:Y)、当該コマンド群をキュー801の先頭に入れる(S905)。

【0082】次に、帯域保証付きのリードコマンドがあるかどうか調査する(S906)。その結果、当該コマンドがある場合は(S906:Y)、当該コマンド群をキュー801の続きに入れる(S907)。

【0083】残ったコマンド群に対しては、上述の性能パラメータに含まれているセクタアドレス変換式を用いて、コマンドのLBAからアクセス対象となる物理シリンダアドレスを算出し、コマンド実行間のシリンダ移動距離、即ちシーク距離が最短になるようにコマンド順序を並べ替えて、キュー801に入れる(S908)。

【0084】次に、性能パラメータを用いて帯域保証付きライトコマンド群に対する実行時間を予測し(S909)、帯域の保証が可能であるか否かを判定する(S910)。その結果、保証できない場合(S910:N)、即ち、ハードディスク装置101の処理能力が要求された帯域に追いつかない場合は、エラー処理を行う(S917)。エラー処理においては、当該コマンドの実行を事前に中止し、ユーザに通知するか、あるいは、情報発信デバイス(例えば、ホスト装置113に接続されたビデオカメラ等)に対して実効転送速度の変更要求(スローダウン)を行う。

【0085】一方、帯域保証が可能である場合は(S910:Y)、上述のコマンド実行時間の予測結果から、ハードディスク装置101のディスクRAM105が、ホスト装置113からの転送データによってフル状態となるコマンドを検出する(S911)。そして、コマンドキュー801内に、ディスクRAM105上のデータにヒットするリードコマンドがあれば、当該リードコマンドを、検出されたコマンドの後に挿入する(S912)。この場合、ディスクRAM105上に蓄えられた転送データがHDA部108に転送されてアンダーランが発生しなければ、リードヒットするコマンドは、複数個挿入しても良い。

【0086】次に、同様に、性能パラメータを用いて帯*

$$T = T_s + T_r + (60/X_s) \times (n/S) + (512 \times 1) / X_r \cdots (3)$$

3)

シーク時間: T_s [秒]

回転待ち時間: T_r [秒]

スピンドル回転速度: X_s [rpm]

* 帯域保証付きリードコマンド群に対する実行時間を予測し(S913)、帯域の保証が可能であるか否かを判定する(S914)。その結果、保証できない場合は(S914:N)、上述のライト時と同様に、当該コマンドの実行を中止する等のエラー処理を行う(S917)。一方、保証可能である場合は(S914:Y)、上述のコマンド実行時間の予測結果からハードディスク装置101のディスクRAM105がHDA部108からの転送データによってフル状態となるコマンドを検出する(S915)。そして、コマンドキュー801内のリードコマンドであってディスクRAM105上のデータにヒットするリードコマンドがあれば、当該リードコマンドを検出されたリードコマンドの後に挿入する(S916)。この場合、ディスクRAM105上に蓄えられた転送データがホスト装置113に転送されてアンダーランが発生しなければ、前述のリードヒットするコマンドは複数個挿入しても良い。

【0087】以上のようにして、コマンドの並び替えが行われる。

【0088】次に、上述のコマンド実行時間の具体的な予測手法について説明する。コマンド実行時間は以下の3つの場合に分けて算出される。

(a) ライトコマンドの場合

ハードディスク装置101はライトキャッシュ制御方式を採用しているため、ホストインタフェースコントローラ117からディスクRAM105へ要求データが転送された時点でコマンド終了とみなされる。従って、転送セクタ数を n 、上述のライト時のホスト実効転送速度を X_w 、1セクタ=512バイトとすると、コマンド実行時間 T は、

$$T = (512 \times n) / X_w \cdots (1)$$

となる。

(b) リードコマンドでかつキャッシュヒットする場合
(a)と同様に、ディスクRAM105からホストインタフェースコントローラ117へ要求データが転送された時点でコマンド終了とみなされる。よって、リード時のホスト実効転送速度を X_r とすれば、コマンド実行時間 T は、

$$T = (512 \times n) / X_r \cdots (2)$$

となる。

(c) リードコマンドでかつキャッシュミスする場合
本ケースでは、記録媒体から要求データを読み出す必要があるため、コマンド実行時間 T は、

【0089】

【数1】

トラック当りのセクタ数: S

となる。

【0090】なお、式(3)は、ホスト実効転送速度が

記録媒体からのデータ読み出し速度よりも速い場合を想定しているが、逆の場合は、ホスト側のデータ転送がネックになるので、

$$T = T_s + T_r + (60/X_s) \times (1/S) + (512 \times n) / X_r \dots (3)$$

となる。

【0092】シーク時間 T_s は、上述のセクタアドレス変換式を用いて目的セクタまでのシーク量を算出し、シーク時間算出式から求めることが可能である。また、回転待ち時間は、スピンドル回転速度から平均回転待ち時間を求めても良いし、あるいは、ハードディスク装置101にて待ち時間を計測しても良い。なお、式(3)では、1つのコマンドで複数のトラックにまたがってアクセスする場合を想定していないが、その場合でもシーク時間算出式からトラックチェンジする時間を算出することができるので実行時間の予測は可能である。

【0093】図12は、上述のコマンド並べ替えフローチャートに基づき、コマンドの実行順序を入れ替えた具体例を示す図である。同図(a)は、並べ替え前の様子を示し、同図(b)は、並べ替え後の様子を示す図である。同図において、SWは、シーケンシャルライトを、RWは、ランダムライトを、RRは、ランダムリードを表す。SW等の後の数字は、各コマンドの発行された順番を表す。例えば、SW0、SW1、SW2、SW3は、この順番で発行された4つのシーケンシャルライトを表す。また、ヒットフラグ1001とは、ホスト装置113において、ハードディスク装置101の性能パラメータ及び発行したコマンド発行要求データに基づいて、各コマンドの対象データがハードディスク装置101のディスクRAM105上に存在するか否かを予測した結果を示すものである。

【0094】図10(b)に示すように、同図(a)に示したコマンド群を並び替えた結果、まず、帯域保証付きのライトコマンド(1)が優先される。次に、性能パラメータからコマンド実行時間を予測した結果、コマンドSW2を発行するとディスクRAM105がフル状態となると予測されたので、帯域保証付きのライトコマンドの後に、コマンドRR0、RR2(2)が挿入される。これらのコマンドは、性能パラメータからディスクRAM105上のデータにヒットすると予想されたものである。最後に、シーク距離が最短になるように並べ替えられた、残りコマンド(3)が置かれる。

【0095】以上、本実施形態では、ホスト装置113がハードディスク装置101の性能を決定付ける性能パラメータからコマンド実行時間を予測し、コマンドの実行順序を並べ替えることにより、システムの最適化を図ることが可能である。

《第3実施形態》次に、本発明の第3の実施形態について説明する。

【0096】図13は、本発明を適用したホームネット

*【0091】
【数2】

ワークシステムの構成を示す図である。同図に示すように、本システムは、セットトップボックス1203と、デジタルテレビ1210と、ハードディスク装置101とからなる。各デバイス101、1203、1210は、IEEE1394で接続されている。

【0097】セットトップボックス1203は、チューナ1204と、A/D変換器1205と、復調器1206と、デスクランブラ1207と、1394インタフェイス部1201と、CPU1208と、RAM1209とを備える。

【0098】セットトップボックス1203は、ケーブル放送や衛星放送等、外部から入力される情報をチューナ1204で検出し、A/D変換器1205でデジタル情報に変換した後、復調器1206、デスクランブラ1207、IEEE1394インタフェイス部1201を介して他の1394デバイスに出力する。

【0099】デジタルテレビ1210は、1394インタフェイス部1201と、MPEGデコーダ1211と、ディスプレイ制御部1212と、ディスプレイ部1213とを備える。

【0100】デジタルテレビ1210は、1394インタフェイス部1201を介して入力した情報をMPEGデコーダ1211でデコードし、ディスプレイ制御部1212を介してディスプレイ部1213で表示する。

【0101】ハードディスク装置101は、図1、図7に示したものと基本的には変わらないが、ディスクインタフェイスコントローラ103のかわりに、IEEE1394インタフェイスを制御するため1394インタフェイス部1201を備えている。

【0102】本実施形態では、ハードディスク装置101等のネットワークに接続された1394デバイスの性能パラメータを、他の1394デバイスに通知する一手法として、1394インタフェイスの初期化時に送受信されるセルフIDパケットを利用する。

【0103】図14は、1394インタフェイスにおける初期化のフローチャートを示す図である。

【0104】1394デバイスのパワーオン/オフやプラグイン/アウトに伴いバスリセットが発生すると、バスの初期化が行われて、すべてのトポロジー情報が消去される(S1302)。次に、新しいトポロジー情報を得るためツリーIDプロセスが実行される(S1303)。ここでは、デバイス間の親子関係とルートとなるデバイスを決定する。最後に、セルフIDパケットを送受信してデバイスの物理IDを含むデバイス情報をお互いに交換する(S1304)。

【0105】図15は、セルフIDパケットの構成を示す図である。本実施形態では、ハードディスク装置101の性能パラメータを他の1394デバイスに通知するために、パケット番号nのパケットを、性能パラメータ情報を通知するためのパケットとしている。そして、このパケットのビット0の中身が「1」であれば、デバイスが性能パラメータを通知する機能を持っていると規定しておく。また、ビット1～31には、性能パラメータが格納されているIEEE1212 CSR（コントロール&ステータスレジスタ）アーキテクチャに準拠したアドレス空間におけるアドレスオフセット値1401を入れる。なお、ビット1～31等に、性能パラメータ値を入れるようにしても良い。

【0106】IEEE1394インタフェースでは、IEEE1212で規定された64ビットのアドレス空間をサポートしている。図16は、IEEE1212で規定された64ビットのアドレス空間を示す図である。IEEE1212規格では、64ビットのアドレスの上位10ビットでバスIDが指定され、続く6ビットでノードIDが指定され、残りの48ビットで各ノードにおけるアドレスが指定される。

【0107】ハードディスク装置101は、自己の性能パラメータ情報を、例えば、イニシャルユニットスペースに配置しておき、配置したアドレスのオフセット1401を上記セルフIDパケット番号nのビット1～31に入れる。ハードディスク装置101が送信したセルフIDパケットを受信したデバイスは、ハードディスク装置101に対してリードブロックリクエストパケットを送信して、オフセットアドレス1401に格納された性能パラメータを読み出す。

【0108】図13に示したシステムでは、ハードディスク装置101から送信されたセルフIDパケットを、セットトップボックス1203及びデジタルテレビ1210が受信する。これらのデバイスは、このセルフIDパケットを受信することにより、ハードディスク装置101が性能パラメータを通知する機能を持っていることを把握でき、更に、性能パラメータを受信することもできる。

【0109】次に、本実施形態におけるコマンドの並べ替えについて説明する。

【0110】本実施形態では、ハードディスク装置101をアクセスするデバイスが複数存在するので、コマンドの並べ替えは、ハードディスク装置101に対してコマンドを発行する複数のデバイスの中からマスターデバイスを選択し、マスターデバイスにおいて行う。本実施形態では、セットトップボックス1203が、マスターデバイスの役割を受け持つとする。なお、ネットワークにつながったデバイスであれば、1394インタフェースの持つデバイス対等（peer to peer）の性質から同様の制御が可能である。

【0111】1394インタフェースは、バスの初期化の際、アイソクロナス転送の管理を行うアイソクロナスマネージャを選択する機能を持つ。アイソクロナスマネージャは、動画等の時間保証を必要とするパケット送信の転送帯域を保証するための管理を行う。送信したいデバイスは、アイソクロナスマネージャに必要とする帯域幅を通知し、許可を受けた後、アイソクロナスパケットを送信する。本実施形態では、アイソクロナスパケットマネージャに選ばれたデバイスが、マスターデバイスの役割を持つようにする。

【0112】ハードディスク装置101にアクセスしたいマスターデバイス以外のデバイスは、マスターデバイスにその旨を通知し、その処理に必要なコマンド群を送信する。これは、通常のアシンクロナスパケットを用いて行えばよい。

【0113】例えば、デジタルテレビ1210がハードディスク装置101にアクセスしたい場合、デジタルテレビ1201は、セットトップボックス1203を経由してハードディスク装置101をアクセスすることになる。まず、デジタルテレビ1201は、セットトップボックス1203に対して、ある特定のアドレス空間にパケットを送付する。この特定のアドレス空間は、予め、マスターデバイスを経由したパケット発行依頼に使用することをすべての1394デバイスに通知しておく。そして、このパケットのデータ部に、パケットの発行依頼者（この場合、デジタルテレビ1201）、パケット転送先（この場合、ハードディスク装置101）、及び、コマンドの内容を格納する。このパケットを受け取ったマスターデバイスであるセットトップボックス1203は、パケット転送先に対してパケット発行依頼者からの要求を送信する。パケットの送受信における確認処理については、1394インタフェースで規定しているように、アシンクロナス転送時は、確認応答処理を行い、アイソクロナス転送時は、確認処理は行わないようにすればよい。

【0114】セットトップボックス1203は、送信されたコマンド群を受けて、他のデバイスから受けたコマンド群と発行順序を調整し、システムパフォーマンスの最適化を図ることができる。

【0115】図17は、図12と同様にコマンドの並べ替えの例を示す図である。ここでは、ハードディスク装置101から動画を読み出してデジタルテレビ1210に出力する処理と、セットトップ1203からハードディスク装置101に対するランダムデータの書き込み及び読み出し処理を同時に行う場合を想定している。

【0116】図17に示すように、デジタルテレビ1201がハードディスク装置101に対して発行する帯域保証付きのシーケンシャルリードコマンド（SRx）と、セットトップボックス1203からのランダムアクセス（RRx、RWx）の要求が来ている。

【0117】並べ替え前のコマンドキュー801の先頭はランダムリードであるが、デジタルテレビ1210への動画再生レスポンスをはやくするために、SR0を最優先させ、かつ、先読み（HDA部108からディスクRAM105へのデータ転送）を実行させる（1）。SR0は、ディスクRAM105上に対象データがなくHDA部108からの読み出しとなるためシーク等に時間がかかる。よって、次に、RR0、RR1を実行する

（2）。これはキャッシュヒットしておりディスクRAM105からセットトップボックス1203へのデータ転送になるので、前記先読みは継続される。次に、SR1～3を実行する（3）。SR1は転送対象データが既にディスクRAM105に格納されているためリードヒットとなる。また、SR2、SR3によるデータ転送によりデジタルテレビ1210側の1394インタフェース部1201のRAMがフル状態となるため、SR4の前にRW0、RR2（4）を実行しても動画が途切れることなく再生することができる。

【0118】以上では、ハードディスク装置101の性能パラメータをセットトップボックス1203に送り、ハードディスク装置101の性能パラメータに基づいて、ハードディスク装置101に発行されるコマンドの実行時間の予測を行ったが、更に、デジタルテレビ1210の性能パラメータを、セットトップボックス1203に通知するようにしてもよい。

【0119】この場合、セットトップボックス1203は、ハードディスク装置101に対して発行するコマンドの実行時間をより正確に予測することが可能となる。例えば、デジタルテレビ1210の1394インタフェース部1201のRAM1202の容量、RAM1202からMPEGデコーダ1211へのデータ出力スピード、RAM1202へのデータ入力スピードを通知すれば、セットトップボックス1203において、RAM1202へのデータ転送によって、いつRAM1202がフルになるかを予測することができる。

【0120】以上説明したように、本実施形態では、ハードディスク装置101をアクセスするデバイスが複数の場合でも、ハードディスク装置101に対してコマンドを発行する複数のデバイスの中からマスターデバイスを選択し、当該マスターデバイスがハードディスク装置101に対するアクセスを一括管理して、コマンド実行時間を予測しコマンドの実行順序を並べ替えることにより、システムの最適化を図ることが可能となる。

【0121】また、動画転送などの帯域保証を必要とするパケット転送においても、IEEE1394バス上の帯域保証だけでなく、ハードディスク装置の内部状態等のシステム状態を考慮したかたちで、データ転送の帯域が保証され、システムの最適化を図ることが可能である。

【0122】なお、第2、第3の実施形態で示したハー

ドディスク装置101は、第1の実施形態と同様に、性能パラメータが可変なものを利用してもよい。この場合、ホスト装置等は、変更後の性能パラメータ値を保存しておき、変更後の性能パラメータ値を使って、コマンド実行時間の予測等を行うようにすればよい。

【0123】

【発明の効果】以上詳細に説明したように、本発明では、外部記憶装置の性能を決定付ける性能パラメータに適切な値を設定することにより、外部記憶装置を含むシステムの最適化を図ることが可能である。

【0124】また、外部記憶装置等の性能パラメータからコマンド実行時間を予測しコマンドの実行順序を並べ替えることにより、システムパフォーマンスの向上を図ることが可能となる。

【図面の簡単な説明】

【図1】 本発明による情報処理システムのブロック図である。

【図2】 ハードディスク装置の性能パラメータの例を示す図である。

【図3】 動作モードの例を示す図である。

【図4】 性能パラメータの設定処理を示すフローチャートである。

【図5】 Identify Device情報の構成を示す図である。

【図6】 アプリケーション実行時の性能パラメータの設定処理を示すフローチャートである。

【図7】 本発明による第2の情報処理システムのブロック図である。

【図8】 コマンド実行処理を示すフローチャートである。

【図9】 コマンドキューの構成を示す図である。

【図10】 コマンド並べ替え処理を示すフローチャート（その1）である。

【図11】 コマンド並べ替え処理を示すフローチャート（その2）である。

【図12】 コマンド並べ替えの一例を示す図である。

【図13】 本発明によるホームネットワークシステムのブロック図である。

【図14】 IEEE1394インタフェースの初期化処理のフローチャートを示す図である。

【図15】 セルフIDパケットのフォーマット例を示す図である。

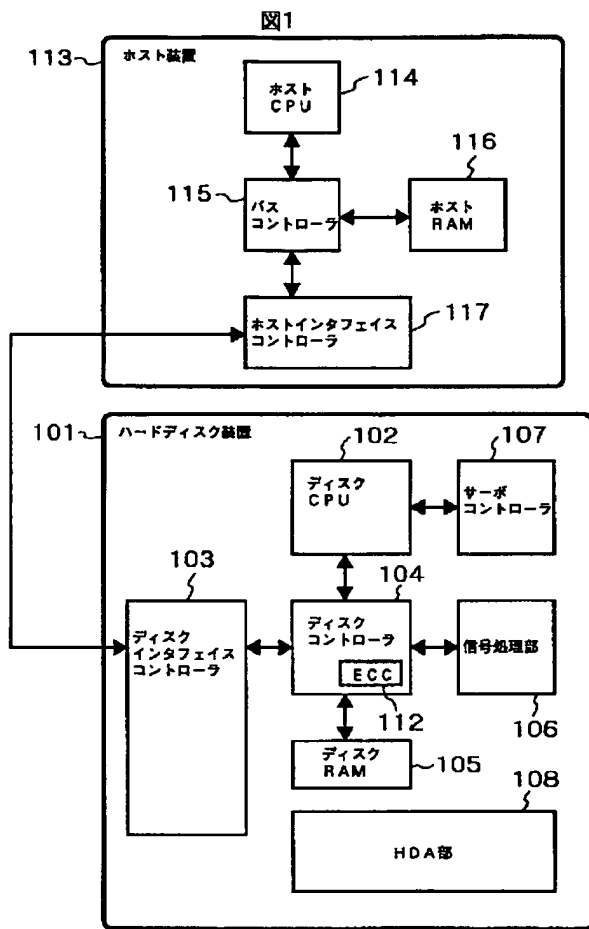
【図16】 IEEE1394インタフェースで使用するアドレス空間の概念図である。

【図17】 コマンド並べ替えの一例を示す図である。

【符号の説明】

101…ハードディスク装置、103…ディスクインタフェースコントローラ、113…ホスト装置、801…コマンドキュー、802…帯域保証フラグ、1201…1394インタフェース部、1203…セットトップボ

【図1】



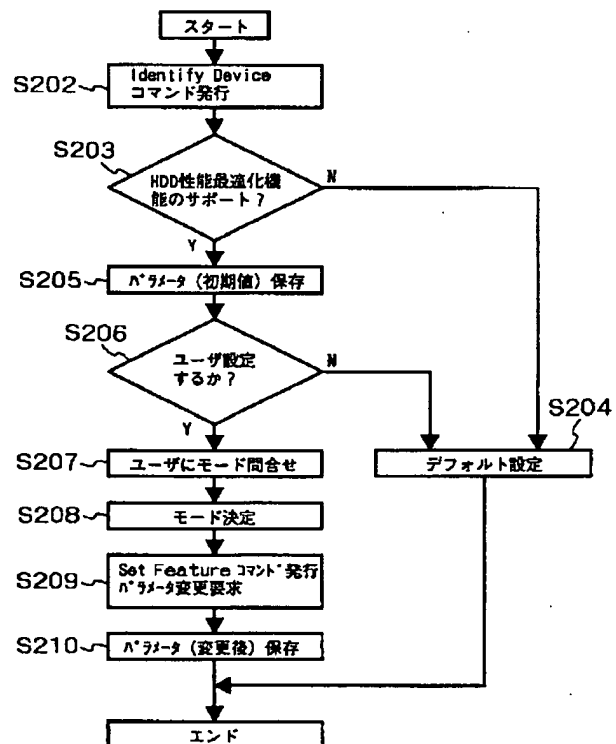
【図2】

図2

オフセットアドレス	パラメータの種類	初期値	設定可能範囲
0	スピンドル回転速度	7200 rpm	3600 / 5400 / 7200
1	キャッシュ制御モード	1	0: 固定セグメント方式 1: 可変セグメント方式
2	キャッシュサイズ	4.0 MB	0.5 / 1.0 / 2.0 / 4.0
3	最大先読みセクタ数	512 セクタ	64 / 128 / 256 / 512
4	シークモード	1	0: 低速モード 1: 高速モード
5	予備セクタ配置方式	0	0: トラック単位 1: シリンダ単位 2: ゾーン単位 3: 磁置単位
6	予備セクタ数	1	1 ~ n
7	ECC長	40 B	10 / 20 / 30 / 40
8	オンザフライECC訂正数	4 箇所	1 / 2 / 3 / 4
9	ホスト転送速度	100 MB/s	16 / 33 / 66 / 100

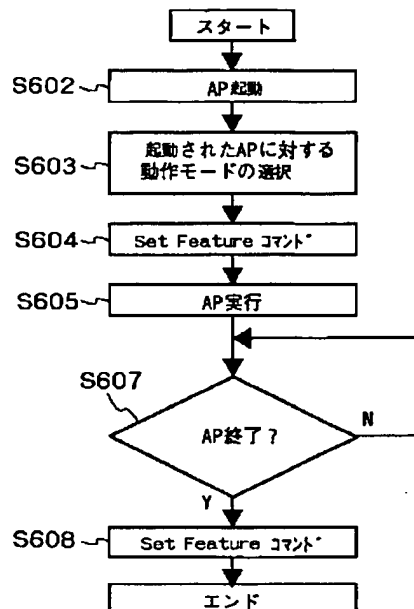
【図4】

図4



【図6】

図6



【図3】

図3

設定値	モード名	オフセット・レス	パラメータ設定例
aaa	ハイ・パフォーマンスモード 501	0	7200
		1	1
		2	4.0
		3	512
		4	1
		5	0
		6	1
		7	40
		8	4
		9	100
bbb	低騒音/低消費電力モード 502	0	3600
		1	1
		2	4.0
		3	512
		4	0
		5	0
		6	1
		7	40
		8	4
		9	16
ccc	動画再生モード 503	0	5400
		1	0
		2	4.0
		3	512
		4	1
		5	2
		6	10
		7	40
		8	4
		9	66
ddd	高信頼モード 504	0	3600
		1	0
		2	1.0
		3	512
		4	0
		5	0
		6	2
		7	40
		8	1
		9	16

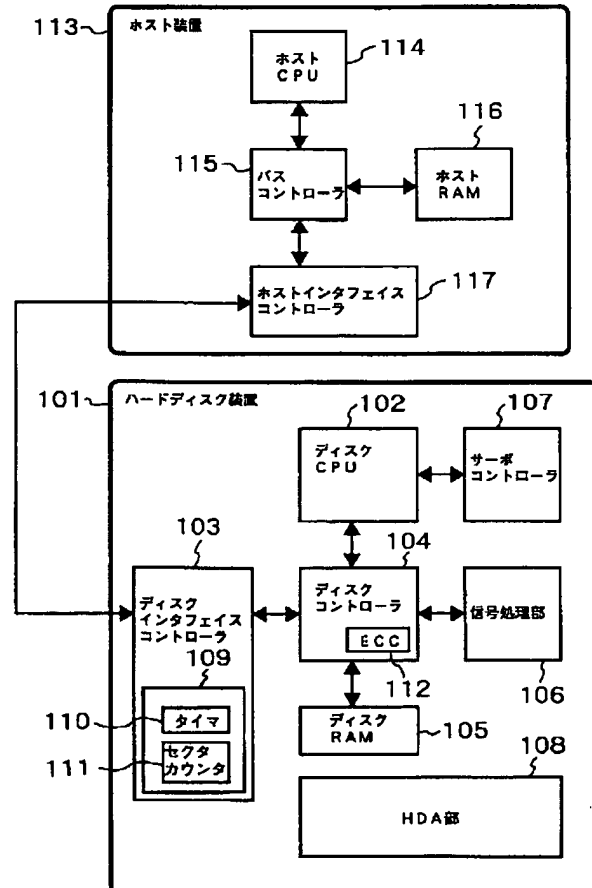
【図5】

図5

Word	Description
0	General Configuration
1	Number of logical cylinders
2	Reserved
...	...
x	bit 15-1 : 性能パラメータ格納ベースアドレス (y) bit 0 : 1: HDD性能最適化機能をサポート 0: HDD性能最適化機能を未サポート
y	性能パラメータ a
y+1	性能パラメータ b
...	...

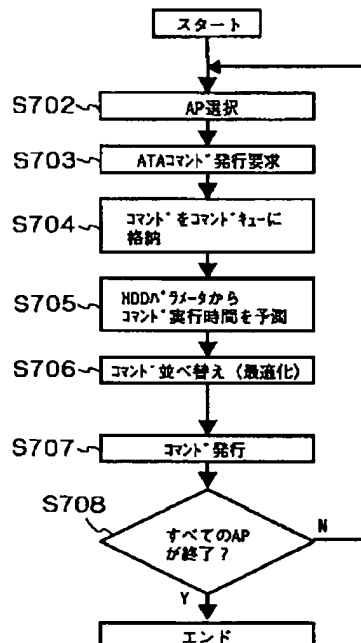
【図7】

図7

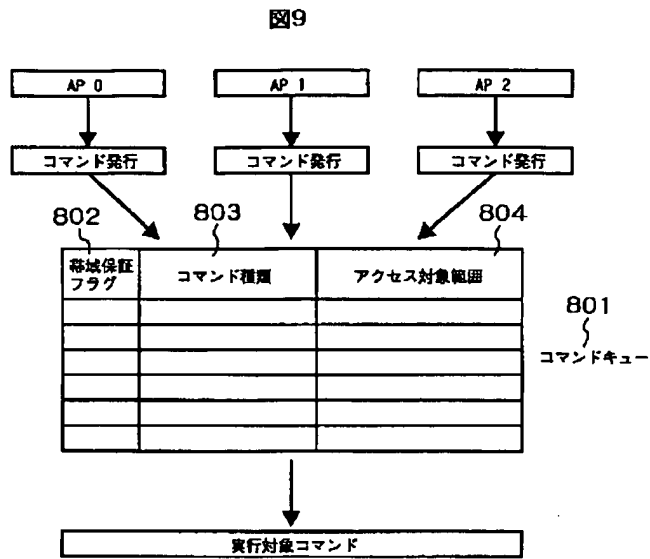


【図8】

図8



【図9】



【図12】

図12

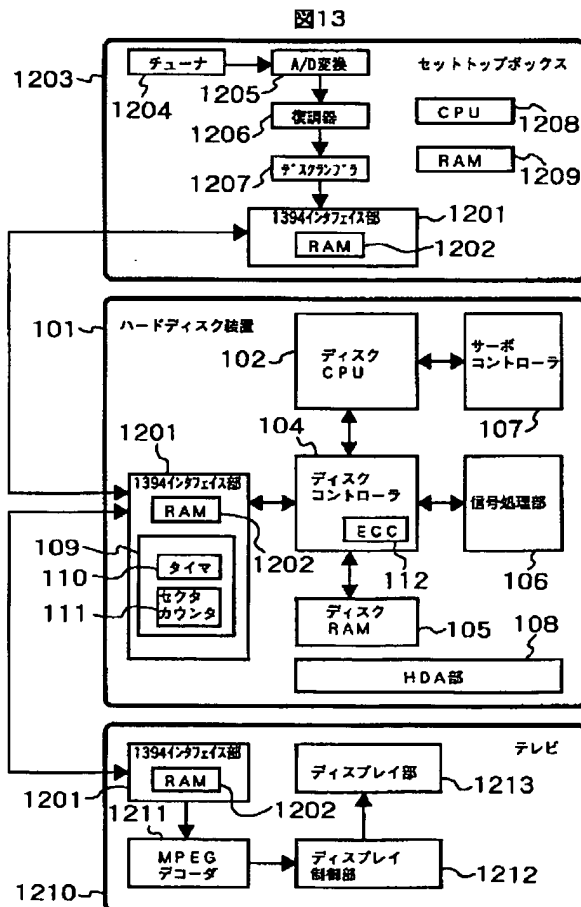
802	803	1001	
帯域保証 フラグ	コマンド	ヒット フラグ	
0	RW 2	0	
1	SW 3	0	
1	SW 2	0	
0	RW 1	0	
0	RW 0	0	
1	SW 1	0	
0	RR 2	1	
0	RR 1	0	
0	RR 0	1	
1	SW 0	1	(1)
1	SW 0	0	

(a) 並べ替え前

802	803	1001	
帯域保証 フラグ	コマンド	ヒット フラグ	
0	RW 2	0	
0	RW 1	0	
0	RW 0	0	
0	RR 1	0	
1	SW 3	0	(3)
0	RR 2	1	
0	RR 0	1	(2)
1	SW 2	0	
1	SW 1	0	
1	SW 0	0	

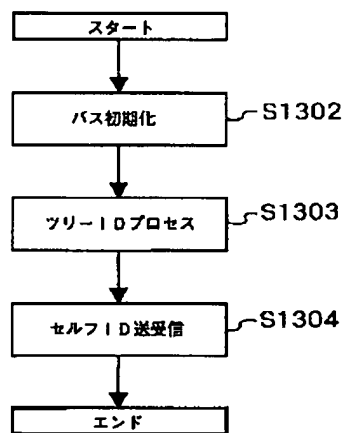
(b) 並べ替え後

【図13】



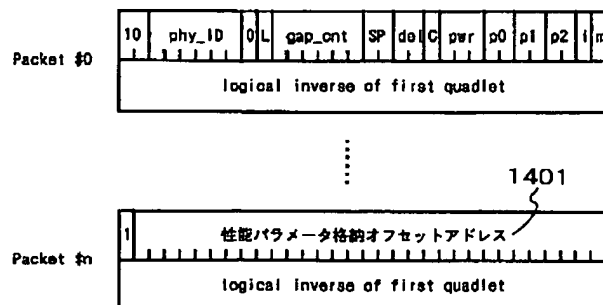
【図14】

図14



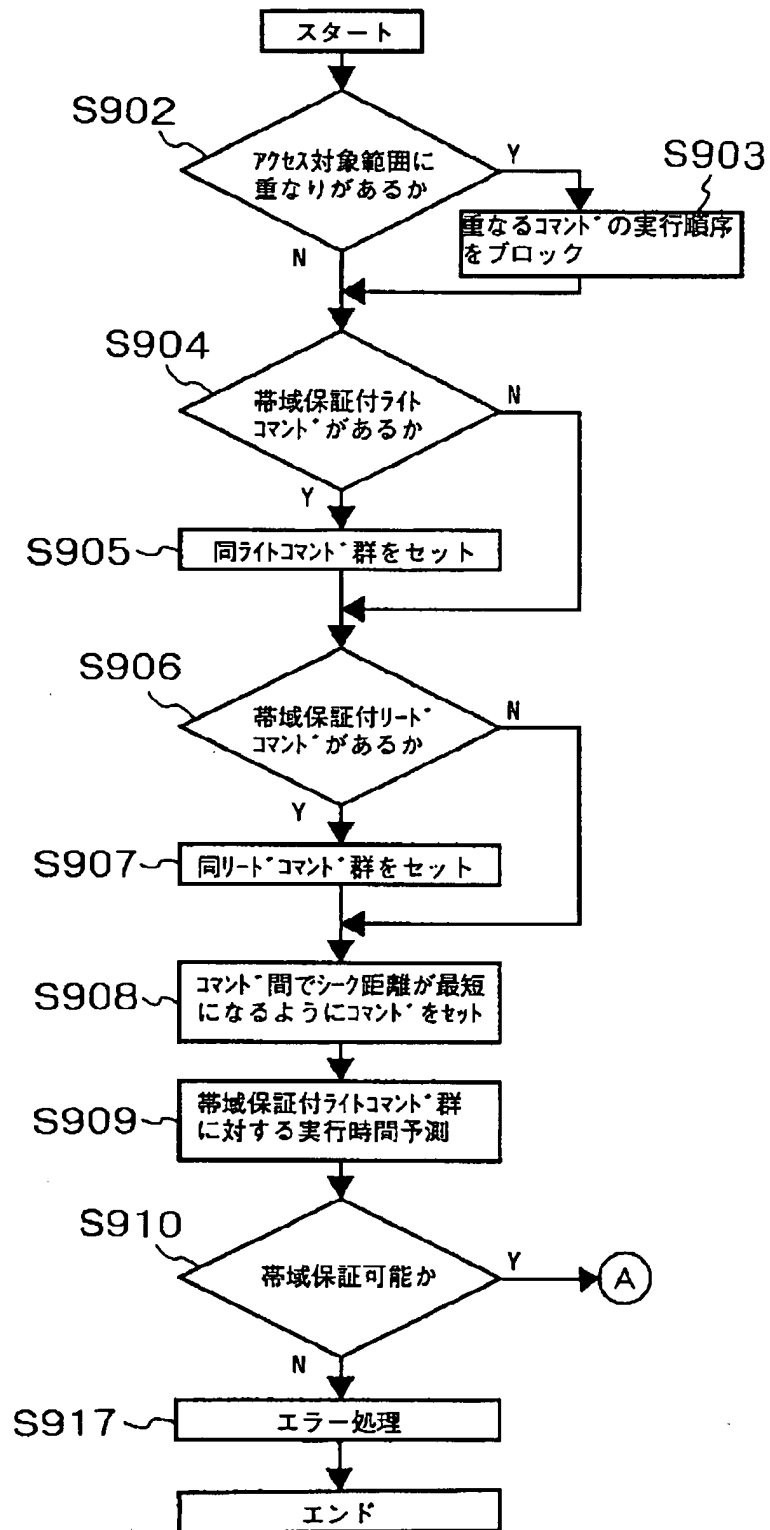
【図15】

図15



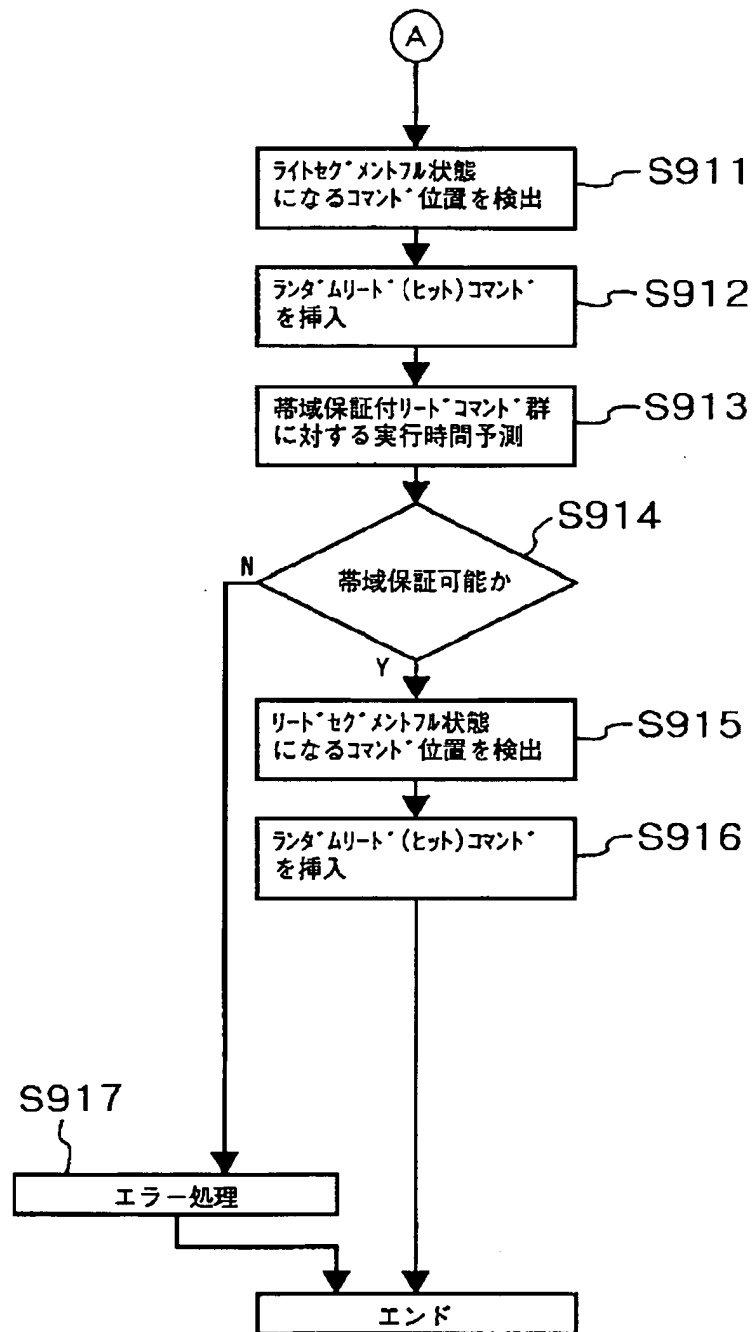
【図10】

図10



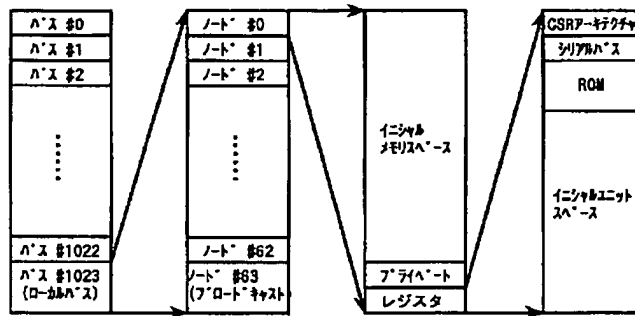
【図11】

図11



【図16】

図16



【図17】

図17

802	803	1001
帯域保証 フラグ	コマンド	ヒット フラグ
0	RR 2	0
0	RR 0	0
0	RR 1	1
1	SR 5	0
1	SR 4	0
1	SR 3	0
1	SR 2	0
1	SR 1	0
1	SR 0	0
0	RR 0	1

(a) 並べ替え前

802	803	1001
帯域保証 フラグ	コマンド	ヒット フラグ
1	SR 5	0
1	SR 4	0
0	RR 2	0
0	RR 0	0
1	SR 3	0
1	SR 2	0
1	SR 1	1
0	RR 1	1
0	RR 0	1
1	SR 0	0

(b) 並べ替え後

フロントページの続き

(72)発明者 井口 慎也
 神奈川県川崎市麻生区王禅寺1099番地 株
 式会社日立製作所システム開発研究所内

(72)発明者 西川 学
 神奈川県小田原市国府津2880番地 株式会
 社日立製作所ストレージシステム事業部内
 F ターム(参考) 5B065 BA01 ZA05
 5D044 AB05 AB07 BC01 CC04 DE42
 GK12 HL01